Predicting Our Future Feelings: The Role of Working Memory for Emotion

by

Colleen C. Frank

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2021

Doctoral Committee:

Professor Patricia A. Reuter-Lorenz, Chair
Professor John Jonides
Professor Ethan Kross
Associate Professor Chandra Sripada

Colleen C. Frank

ccfrank@umich.edu

ORCID iD: 0000-0003-4256-8983

Dedication

This dissertation is dedicated to my incredible parents—your unwavering support and enthusiastic encouragement has undoubtably gotten me to where I am today.

Acknowledgements

First and foremost, I would like to thank my wonderful advisor, Patti Reuter-Lorenz, for all of your support over the last five years. You have positively impacted my professional and personal growth in innumerable ways but above all else, I am so appreciative of how you have treated me. During my graduate career, you have consistently made me feel as though I was a respected and highly-valued member of our team with worthwhile ideas and contributions. Thank you for making every effort to ensure that I knew that I was a priority, and that you cared, not just about my development as a scholar, but about me as an individual. You have been a truly remarkable mentor and I am confident that I would not be the scientist I am today if it weren't for your patience, kindness, and guidance.

Thank you to other members of my committee—John Jonides, Ethan Kross, and Chandra Sripada—for the thoughtful insight you provided throughout the entire dissertation process. Thank you to my collaborators and coauthors on the manuscript from Chapter 2—Alex Iordan and Joseph Mikels, for your support and feedback on this project. And thank you to the many dedicated research assistants that I have had the privilege of working with over the last few years. Mentoring has been one of my true joys of graduate school and it has been so exciting to watch all of you flourish.

Thank you to Lilian Cabrera-Haro, my cohort-mate, lab-mate, and dear friend, for being by my side every single step of the way. You were nothing short of the perfect partner to tackle graduate school with and I appreciate you so much. Thank you to Kathy Xie for being a fantastic lab-mate, colleague, and friend, and for being the best hype woman I know. Thank you to my fellow psychology department graduate students, especially Andrea Belgrade, Cody Cao, Kristi Chin, Stella Hao, and Selena Tran, who, in these final few months, have provided me with the emotional support, food, gifs, and walks I needed to get across the finish line. And thank you to members of the LIFE program for all of your feedback and support over the years. I feel so lucky to have had the opportunity to establish several meaningful and lasting friendships with fellows from around the world.

To those I consider part of my Michigan family—Tessa Abagis, Lilian Cabrera-Haro, Joe DesOrmeaux, Myrna Cintron-Valentin, Josh Haro, Tiffany Jantz-DesOrmeaux, Mike Kling, and Jordan Meyer—thank you for all of the fun and the wonderful memories. Whether it was an impromptu party for bad news, a themed potluck, or the countless hours of karaoke, your friendship has meant the world to me and I cannot wait for our next non-virtual reunion.

I am forever indebted to my parents—Lawrence and Theresa Frank, siblings—Conor Frank and Caitrin Sobota, and closest friends—Jenny Hinz, Matt Caudill, and Megan Yung, for accompanying me on this journey that none of you signed up for. I appreciated every single phone call, text, email, gif, meme, video and photo that reminded me that you were there rooting for me. Thank you for being my personal cheerleaders and professional pep talkers for years on end.

And finally, thank you to Tara Lineweaver, a psychology professor at my alma mater, Butler University, and the woman who inspired me—a 17-year-old, prospective college student—to obtain my PhD. I will forever be grateful for the path you led me on and for the role you played in getting me here.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

One key element of decision making is predicting how different outcomes will make us feel. While individuals vary considerably in this ability, known as affective forecasting, the reasons for this variability in prediction accuracy are largely unknown. This dissertation aims to uncover why some people are better forecasters than others. We test the hypothesis that affective forecasting is supported by a fundamental psychological ability—i.e., *affective working memory*. Affective working memory refers to the distinct domain of working memory responsible for actively maintaining and working with feeling states. Because affective forecasting requires conjuring up and maintaining emotional experiences for evaluation, we predicted that individuals who are better able to maintain feeling states (i.e., superior affective working memory), would be more accurate in predicting their future feelings. Across Study 1 (Study 1a: N = 66; Study 1b: N = 68) and Study 2 (N = 96), results emerged in support of this hypothesis such that working memory for emotion uniquely predicted affective forecasting accuracy, whereas working memory for perceptual (i.e., brightness) information, did not. In Study 3 (N = 85), we more firmly establish the selective relationship between forecasting accuracy and affective working memory, finding that performance on two additional, widely used measures of visual working memory does not explain any variability in forecasting accuracy beyond the variability explained by affective working memory. In Study 4 (N =76), we find that this relationship between affective working memory and forecasting accuracy generalizes to a real-world event where we measured predicted and experienced feelings to the outcome of the 2020 United States presidential election. Thus, across the studies presented in this dissertation, we find consistent evidence that individual differences in the ability to maintain emotional experiences predicts variability in forecasting accuracy. These findings advance our understanding of the mechanistic underpinnings of forecasting accuracy and demonstrate that affective working memory is a core psychological process underlying higher-order emotional thought, including emotional prospection.

Chapter 1 Introduction

A crucial aspect of making good decisions is predicting how different outcomes will make us feel. From deciding which job to take, which house to buy, or who to marry, the ability to predict our future feelings is fundamental to wellbeing. This ability, known as *affective forecasting* (Wilson & Gilbert, 2003), varies greatly among individuals, yet the reasons for this variability are poorly understood. This dissertation aims to uncover why some people are better forecasters than others. While previous studies have proposed that forecasting accuracy may be explained by individual differences in trait-level abilities (i.e., emotional intelligence, personality traits), we hypothesize that affective forecasting relies specifically on *affective working memory*—a distinct domain of working memory responsible for actively maintaining and working with feeling states. Because affective forecasting requires conjuring up and maintaining emotional experiences for evaluation, we predicted that individuals who are better able to maintain feeling states (i.e., better affective working memory), would be more accurate in predicting their future feelings. In contrast, we predicted that the ability to maintain non-affective (i.e., cognitive) information would not account for variability in forecasting. Evidence in support of these hypotheses would lend credence to the idea that affective working memory is a unique, fundamental capacity that contributes to emotional prospective thought. Such evidence would also uncover a key component of affective forecasting—an ability that has yet to be decomposed into psychological subprocesses.

Affective Forecasting

We have known for centuries that feelings play a role in decision making (Bentham, 1789; Jevons, 1879; Edgeworth, 1881). However, it wasn't until more recently that researchers began to incorporate emotion into theoretical models of choice behavior. One of the primary ways that emotions influence our decisions is through *affective forecasting*, i.e., predicting our future feelings. According to Decision Affect Theory (DAT; Mellers, Schwartz, Ho & Ritov, 1997), when making a decision, we consider the anticipated feelings for each outcome to ultimately guide us to our choice. This idea is supported by robust evidence that models of decision making that include expected feelings predict choice behavior above and beyond those that do not account for anticipated emotions (Mellers, Schwartz, & Ritov, 1999; Charpentier et al., 2016; Ahn et al., 2012; Hayes & Wedell, 2020). Moreover, not only can anticipated emotions be used to predict our choices, but inaccuracies in these forecasts may lead to suboptimal financial and medical decision making (Kermer et al., 2006; Loewenstein et al., 2001; Dunn & Laham, 2006; but see Sharot, 2011). Thus, the ability to accurately predict future feelings is essential to making optimal decisions.

Forecasting abilities are typically assessed by having participants first predict how they would feel if a given event or outcome were to occur. Once the event occurs, participants report their experienced feelings. Forecasting accuracy can then be calculated as the difference between ratings of predicted and experienced feelings, where a larger discrepancy indicates a less accurate prediction and thus, poorer forecasting ability. Participants may be asked to predict one or more aspects of their future feelings including the valence, specific emotions, intensity, and duration (Wilson & Gilbert, 2003). Depending on the study, participants may anticipate their feelings to personal life events (e.g., attaining tenure, entering/exiting romantic relationships), major national or world events (e.g., election outcomes), or experimental tasks designed to simulate processes that are also at play during real-world forecasting (e.g., viewing emotionally-evocative scenes).

Despite affective forecasting's importance, such predictions are not always accurate. Whereas forecasts of valence and the identification of future emotions are fairly precise (e.g., correctly predicting you'll feel happy if your beloved sports team wins), predictions of how intense, and how long feelings will last, are more prone to error (Wilson & Gilbert, 2003). In general, there is a tendency for individuals to overestimate the intensity and duration of their feelings, known as the impact bias (Gilbert et al., 1998; Wilson et al., 2000; Wilson & Gilbert, 2003). Some experts have claimed that the extent of the impact bias has been overstated, likely as a result of vague wording of the prompts used at the time of prediction (Levine et al., 2012; Lench et al., 2019). However, others continue to endorse the importance of impact bias (Wilson & Gilbert, 2013) and have proposed several explanations that can account for inaccurate forecasting (Wilson & Gilbert, 2003; 2005). For example, when predicting future feelings, people may overestimate the impact of the target event while minimizing the influence of concurrent events (i.e., *focalism*; Schkade & Kahneman, 1998; Wilson et al., 2000). Other accounts of forecasting error include the idea that people may ignore the mind's tendency to rationalize the effectiveness of the so-called 'psychological immune system' that is responsible for reducing levels of negative affect (i.e., *immune neglect*; Gilbert et al., 1998; Halpern & Arnold, 2008). These misestimations may be further exacerbated by differences between the affective states at the time of prediction and at the time of experience (i.e., *hot-cold empathy gap*; Loewenstein, 2005). Notably, these accounts of inaccurate predictions are not mutually exclusive and it is likely that several factors contribute to errors in forecasting.

Given the multiple potential sources of error, it is unsurprising that people vary considerably in their forecasting accuracy (Dunn et al., 2007; Hoerger et al., 2012a; Hoerger et al., 2016). That is, the degree of misestimation is not uniform across individuals, and some people are better forecasters than others. Hence, researchers have also examined individual difference variables that may explain person-to-person variation in forecasting accuracy. Such efforts have focused primarily on psychological traits that could conceivably predispose one to be more or less accurate at forecasting. One appealing idea is related to the notion of emotional intelligence, a construct that refers to the set of processes related to perceiving, understanding, and managing emotion (Mayer, Salovey, & Caruso, 2004; Petrides & Furnham, 2001). Indeed, there is evidence to suggest that emotional intelligence underlies some variability in prediction accuracy, with at least two studies finding that those with higher emotional intelligence are also more accurate predictors of their future feelings (Dunn et al., 2007; Hoerger et al., 2012a). Forecasting accuracy has also been linked to personality traits, such as openness and extraversion (Hoerger et al., 2016), and other studies have found that depressive symptoms are associated with more blunted positive, and more extreme negative, predictions (Strunk, Lopez, & DeRubeis, 2006; Wenze et al., 2012; Hoerger et al., 2012b; Marroquín & Nolen-Hoeksema, 2015).

However, this dissertation takes a different approach to explaining forecasting error by focusing on the nature of the mental workspace in which affective forecasts are formulated. Our perspective is informed by the field of cognitive psychology, where it is well-established that individual differences in higher-order cognitive processes, such as planning and problem-solving, are due, at least in part, to individual differences in cognitive working memory (Jonides, 1995; Engle & Kane, 2003). That is, according to decades of research, variability in working memory can predict performance on reasoning, reading comprehension, and other tasks where information must be kept on-line (Daneman & Carpenter, 1980; Kane et al., 2004; Unsworth et al., 2015). By analogy, we posit that *affective working memory*—a domain-specific working memory subsystem dedicated to the maintenance of feeling states—is crucial for higher-order emotion-based thought, including affective forecasting. If affective working memory provides the mental workspace that supports emotional prospection, then individual differences in the working memory for emotion should account for variability in the accuracy of a persons' forecasts.

Affective Working Memory

Working memory refers to the ability to hold information at the forefront of our minds (Baddeley, 1992; Miyake and Shah, 1999). For example, when we perform mental arithmetic, or consider a seating arrangement for a dinner party, we are using our working memory abilities. We can hold emotions in mind as well (Davidson & Irwin, 1999; Mikels et al., 2005; 2008). For example, when we consider going out to eat or ordering food in, we must keep the associated feelings in mind for comparison in order to make a decision. This ability to maintain and work with feeling states is called *affective working memory* (Mikels & Reuter-Lorenz, 2019). Affective working memory reflects a distinct domain of working memory responsible for actively holding emotional experiences in mind in the service of goal directed thought and behavior.

To measure the ability to maintain subjective experiences of emotion, Mikels and colleagues (2003; 2005; 2008) designed a novel emotion maintenance task. In this task, participants view one emotional image and maintain the feelings evoked by the image over a delay. Another emotional image is then presented and participants must decide if the second image elicited feelings with more or less emotional intensity (i.e., strength or amount of emotional reaction) compared to the first. Other versions of the task remove the comparison to a second image, and participants maintain the elicited feelings before categorizing them as having either high or low emotional intensity (Waugh, Lemus, & Gotlib, 2014; Waugh et al., 2019). In both versions, ratings of emotional intensity from all images are collected separately and then used to determine how participants should have responded during the maintenance task. Higher emotion-maintenance accuracy scores reflect a superior ability to hold emotions in mind, therefore, better affective working memory.

Affective working memory can be characterized by its separability from other mental processes including cognitive working memory, working memory for emotionally-laden stimuli, and the passive experience of emotion. We consider each of these in turn. First, evidence indicates that affective working memory is dissociable from cognitive (i.e., non-affective) working memory abilities. When developing the emotion maintenance task, the authors also created a perceptual maintenance task that assesses the ability to maintain brightness information (Mikels, 2003; Mikels et al., 2005; 2008). This companion task was designed to measure non-affective maintenance in an analogous manner to that of the emotion maintenance task; both maintenance tasks require assessments of the intensity of subjective experiences evoked from pictorial stimuli. In this task, participants hold the brightness intensity (i.e., amount of overall light or illumination) of a neutral image in mind over a delay before deciding if a second neutral image has more or less brightness intensity compared to the first (Mikels et al., 2005, 2008; Broome et al., 2012). The emotion and brightness maintenance tasks have been used in several studies to demonstrate the dissociation between maintenance of affective and non-affective information (Mikels et al., 2005; 2008). Using selective interference methodology, Mikels et al.

(2008) found that concurrent performance of a secondary affective task interrupted emotion, but not brightness-maintenance accuracy, whereas completing a secondary cognitive task impaired brightness-maintenance, but facilitated emotion-maintenance performance. Additionally, Mikels and colleagues (2005) found that while younger adults outperformed older adults on the brightness maintenance task, older adults performed comparably to younger adults on the emotion maintenance task. This finding is consistent with the relative preservation of emotional processing abilities seen in older age despite robust age-related cognitive decline (Mather, 2016). Further support of the separability between cognitive and affective working memory comes from a study reporting that individuals with schizophrenia, who have known emotional processing deficits (Kring & Elis, 2013), show emotion-maintenance impairments above and beyond those seen on the brightness maintenance task (Gard et al., 2011).

According to Mikels & Reuter-Lorenz (2019), affective working memory constitutes a distinct mode of emotion–working memory interaction responsible for the maintenance of emotional *feelings* themselves, rather than emotional stimuli more generally (e.g., emotional photos, words, faces). When maintaining emotional stimuli actively in mind, unless otherwise directed, the focus of attention is on the stimulus content and not necessarily on the emotional feeling state that the words or images may have evoked. That is, while verbal working memory is at play when holding sad emotional words in mind, affective working memory is engaged when maintaining and working with the specific feelings of sadness evoked by the emotionally laden words.

Moreover, affective working memory is dissociable from purely experiencing emotion. Evidence from neuroimaging studies support the idea that emotion maintenance reflects an active process of holding emotional experiences online that is distinct from the passive experience of emotion. For example, emotion maintenance, as compared to passive viewing of emotional images, was shown to increase activation of the dorsal medial frontal cortex (DMFC; Waugh, Lemus, & Goltib, 2014; Smith et al., 2017), part of the 'mentalizing network' associated with emotion appraisal and regulation (Amodio & Frith, 2006), and the dorsal lateral prefrontal cortex (DLPFC; Waugh et al., 2014; Smith et al., 2017), an area associated with working memory and other executive functioning (Smith & Jonides, 1999; D'esposito & Postle, 2015). Additionally, another study found prolonged activation of the amygdala after emotion maintenance that was not seen after simply experiencing the emotions (Schaefer et al., 2002). Further evidence of a distinction between passive experience and active maintenance of emotions comes from behavioral work that used cognitive load manipulations (DeFraine, 2016). In this study, participants reported that concurrent performance of another mental task reduced the intensity of feelings while experiencing, but not maintaining, feeling states. Thus, affective working memory appears to reflect a unique ability to actively hold emotions in mind.

As stated earlier, we know that cognitive working memory supports goal-directed behavior and that variability in working memory can predict higher-order cognitive processing (Kane et al., 2004). Similarly, affective working memory is hypothesized to operate in service of subsequent goals and that individuals differ in their ability to maintain emotional experiences. Thus, affective working memory may play a comparable role as cognitive working memory providing the mental workspace that supports higher-order mental abilities in the affective domain. In this dissertation, we test the idea that affective working memory underlies one such important emotion-related ability—i.e., affective forecasting.

Overview of the Dissertation

In the present dissertation, we investigate affective working memory as a potential mechanism underlying affective forecasting. Because affective forecasting requires conjuring up and maintaining emotional experiences for evaluation, we hypothesized that individuals who are better able to maintain feeling states (i.e., superior affective working memory), would be more accurate in predicting their future feelings. If this hypothesis is correct, then individual differences in the ability to maintain emotional experiences should predict variability in the accuracy of affective forecasts. Evidence of such a relationship could inform future development of an information processing model of affective forecasting that articulates the specific roles of constituent psychological processes, including affective working memory, in accurate prospection.

In Chapter 2, we present findings from a series of three studies (Frank et al., 2021) that test the hypothesis that affective, but not cognitive, working memory supports affective forecasting accuracy. To measure working memory abilities, participants completed the emotion maintenance and brightness maintenance tasks described above (Mikels, 2003; Mikels et al., 2005; 2008; Broome et al., 2012). To measure affective forecasting, participants reported their predicted and experienced feelings to viewing emotionally-evocative photographs in the laboratory. In Study 1, we find that emotion-maintenance performance predicts forecasting accuracy, whereas brightness-maintenance performance does not. In Study 2, we replicate this selective relationship between emotion, but not brightness, maintenance and forecasting accuracy within the same group of participants. And in Study 3, we more firmly establish the finding that forecasting ability is not independently predicted by performance on the brightness maintenance task, or other measures of cognitive working memory. Furthermore, we find no support, across several studies, that emotional intelligence is associated with forecasting accuracy.

In Chapter 3, we present results from one experiment (Study 4) that examines whether the relationship between affective working memory and forecasting accuracy generalizes to a real-world event by measuring predicted and experienced feelings to the outcome of the 2020 United States presidential election. Moreover, unlike the prior studies in this dissertation which were all conducted in-person, with college students run in a laboratory setting, Study 4 was conducted entirely online with participants recruited nationwide, thereby establishing another dimension to the generality of the effects. We found that the selective relationship between emotion, but not brightness, maintenance and forecasting accuracy was further replicated in this broader sample of participants that used forecasted feelings to a real-world event. The theoretical implications from the findings of all studies are integrated and reviewed in Chapter 4.

References

Ahn, W.-Y., Rass, O., Shin, Y.-W., Busemeyer, J. R., Brown, J. W., & O'Donnell, B. F. (2012). Emotion-based reinforcement learning. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*(34), 78–83.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268–277. https://doi.org/10.1038/nrn1884

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. http://dx.doi.org/10.1126/science.1736359

Bentham, J. (1789). *An introduction to the principles of morals and legislation.*

Broome, R., Gard, D. E., & Mikels, J. A. (2012). Test-retest reliability of an emotion maintenance task. *Cognition and Emotion*, *26*(4), 737–747. https://doi.org/10.1080/02699931.2011.613916

Charpentier, C. J., De Neve, J. E., Li, X., Roiser, J. P., & Sharot, T. (2016). Models of affective decision making: How do feelings predict choice? *Psychological Science, 27*(6), 763–775. https://doi.org/10.1177/0956797616634654

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Davidson, R. J., & Irwin, W. M. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences*, *3*, 11–21. https://doi.org/10.1016/S1364-6613(98)01265-0

DeFraine, W. C. (2016). Differential effects of cognitive load on emotion: Emotion maintenance versus passive experience. *Emotion*, *16*(4), 459–467. https://doi.org/10.1037/emo0000140

Dunn, E. W., Brackett, M. A., Ashton-James, C., Schneiderman, E., & Salovey, P. (2007). On emotionally intelligent time travel: Individual differences in affective forecasting ability. *Personality and Social Psychology Bulletin*, *33*(1), 85–93. https://doi.org/10.1177/0146167206294201

Dunn, E. W., & Laham, S. M. (2012). Affective forecasting: A user's guide to emotional time travel. In J. P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 177–194). Psychology Press. https://doi.org/10.4324/9780203720752

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to* *the moral sciences*. Paul, C. K.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), The *psychology of learning and motivation: advances in research and theory*, *Vol. 44* (pp. 145–199). Elsevier Science. https://doi.org/10.1016/S0079-7421(03)44005-X

Frank, C. C., Iordan, A. D., Ballouz, T. L., Mikels, J. A., & Reuter-Lorenz, P. A. (2021). Affective forecasting: A selective relationship with working memory for emotion. *Journal of Experimental Psychology: General, 150*(1), 67–82. https://doi.org/10.1037/xge0000780

Gard, D. E., Cooper, S., Fisher, M., Genevsky, A., Mikels, J. A., & Vinogradov, S. (2011). Evidence for an emotion maintenance deficit in schizophrenia. *Psychiatry Research, 187*(1), 24–29. https://doi.org/10.1016/j.psychres.2010.12.018

Gilbert, D. T., Wilson, T. D., Pinel, E. C., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Halpern, J., & Arnold, R. M. (2008). Affective forecasting: An unrecognized challenge in making serious health decisions. *Journal of General Internal Medicine*, *23*(10), 1708–1712. https://doi.org/10.1007/s11606-008-0719-5

Hayes, W. M., & Wedell, D. H. (2020). Modeling the role of feelings in the Iowa Gambling Task. *Decision*, *7*(1), 67–89. https://doi.org/10.1037/dec0000116

Hoerger, M., Chapman, B., & Duberstein, P. (2016). Realistic affective forecasting: The role of personality. *Cognition and Emotion*, *30*(7), 1304–1316. https://doi.org/10.1080/02699931.2015.1061481

Hoerger, M., Chapman, B. P., Epstein, R. M., & Duberstein, P. R. (2012a). Emotional intelligence: A theoretical framework for individual differences in affective forecasting. *Emotion*, *12*(4), 716–725. https://doi.org/10.1037/a0026724

Hoerger, M., Quirk, S. W., Chapman, B. P., & Duberstein, P. R. (2012b). Affective forecasting and self-rated symptoms of depression, anxiety, and hypomania: Evidence for a dysphoric forecasting bias. *Cognition & Emotion, 26*(6), 1098–1106.

Jevons, W. S. (1879). *The theory of political economy*. Macmillian.

Jonides, J. (1995). Working memory and thinking. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science: Thinking* (2nd ed., pp. 215–265). The MIT Press.

Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science*, *17*(8), 649–653. https://doi.org/10.1111/j.1467-9280.2006.01760.x

Kring, A. M., & Elis, O. (2013). Emotion deficits in people with schizophrenia. *Annual Review of Clinical Psychology, 9*, 409–433. https://doi.org/10.1146/annurev-clinpsy-050212-185538

Lench, H. C., Levine, L. J., Perez, K., Carpenter, Z. K., Carlson, S. J., Bench, S. W., & Wan, Y. (2019). When and why people misestimate future feelings: Identifying strengths and weaknesses in affective forecasting. *Journal of Personality and Social Psychology, 116*(5), 724–742. https://doi.org/10.1037/pspa0000143

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology*, *103*(4), 584–605. https://doi.org/10.1037/a0029544

Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology*, *24*(4), S49–S56. https://doi.org/10.1037/0278-6133.24.4.S49

Loewenstein, G. F., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Marroquín, B., & Nolen-Hoeksema, S. (2015). Event prediction and affective forecasting in depressive cognition: Using emotion as information about the future. *Journal of Social and Clinical Psychology*, *34*(2), 117–134. https://doi.org/10.1521/jscp.2015.34.2.117

Mayer, J., Salovey, P., & Caruso, D. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*, 197–215. http://dx.doi.org/10.1207/s15327965pli1503_02

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8*(6), 423–429. https://doi.org/10.1111/j.1467-9280.1997.tb00455.x

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*(3), 332–345. https://doi.org/10.1037/0096-3445.128.3.332

Mikels, J. A. (2003). *Hold on to that feeling: Working memory and emotion from a cognitive neuroscience perspective* (Doctoral Dissertation). http://hdl.handle.net/2027.42/123450

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging*, *20*(4), 542–553. https://doi.org/10.1037/0882-7974.20.4.542

Mikels, J. A., Reuter-Lorenz, P. A. (2019). Affective working memory: An integrative psychological construct. *Perspectives in Psychological Science*, 1–17. https://doi.org/10.1177/1745691619837597

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion, 8*(2), 256–266. https://doi.org/10.1037/1528-3542.8.2.256

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15,* 425–448. https://doi.org/10.1002/per.416

Schaefer, S. M., Jackson, D. C., Davidson, R. J., Aguirre, G. K., Kimberg, D. Y., & Thompson-Schill, S. L. (2002). Modulation of amygdalar activity by the conscious regulation of negative emotion. *Journal of Cognitive Neuroscience*, *14*(6), 913–921. https://doi.org/10.1162/089892902760191135

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*(5), 340–346. https://doi.org/10.1111/1467-9280.00066

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657–1661. https://doi.org/10.1126/science.283.5408.1657

Smith, R., Lane, R. D., Alkozei, A., Bao, J., Smith, C., Sanova, A., Nettles, M., & Killgore, W. D. S. (2017). Maintaining the feelings of others in working memory is associated with activation of the left anterior insula and left frontal-parietal control network. *Social Cognitive and Affective Neuroscience*, *12*(5), 848–860. https://doi.org/10.1093/scan/nsx011

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy, 44*(6), 861–882. https://doi.org/10.1016/j.brat.2005.07.001

Unsworth, N., Fukunda, K., Awh, E., & Vogel, E. K. (2015). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Waugh, C. E., Lemus, M. G., & Gotlib, I. H. (2014). The role of the medial frontal cortex in the maintenance of emotional states. *Social Cognitive and Affective Neuroscience*, *9*(12), 2001–2009. https://doi.org/10.1093/scan/nsu011

Waugh, C. E., Running, K. E., Reynolds, O. C., & Gotlib, I. H. (2019). People are better at maintaining positive than negative emotional states. *Emotion*, *19*(1), 132–145. https://doi.org/10.1037/emo0000430

Wenze, S. J., Gunthert, K. C., & German, R. E. (2012). Biases in affective forecasting and Recall in individuals With depression and anxiety symptoms. *Personality and Social Psychology Bulletin*, *38*(7), 895–906. https://doi.org/10.1177/0146167212447242

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, *35*(3), 345–411. https://doi.org/10.1016/S0065-2601(03)01006-2

Wilson, T. D., & Gilbert D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology*, *105*(5), 740–748. https://doi.org/10.1037/a0032662

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–836. https://doi.org/10.1037/0022-3514.78.5.821

Chapter 2 A Selective Relationship between Affective Working Memory and Affective Forecasting (Studies 1, 2, & 3)

Our decisions are often guided by the feelings we anticipate in response to potential outcomes. Indeed, the ability to predict one's future feelings, known as affective forecasting (AF) (Wilson & Gilbert, 2003), is thought to play an important role in decisions, preferences, and behaviors that affect health and economic outcomes, and personal well-being. However, people are often biased and inaccurate when predicting their future feelings (Wilson & Gilbert, 2003; Mathieu & Gosling, 2012). Therefore, understanding the elemental processes that contribute to AF may provide valuable insight into why AF is flawed and how it can be improved. We posit that AF entails the ability to actively maintain and evaluate emotional feelings, and thus requires affective working memory (AWM). AWM maintains feeling states actively in mind to support goal-oriented behavior (Mikels, Larkin, Reuter-Lorenz, & Carstensen, 2005; Mikels, Reuter-Lorenz, Beyer, & Fredrickson, 2008; Smith & Lane, 2015; see Mikels & Reuter-Lorenz, 2019, for a review). If AWM is a core ability underlying AF, then individual differences in AWM should influence the accuracy of AF. The goal of the present report is to test this hypothesis.

Affective forecasts typically rely on self-reports in which participants predict how they will feel about future stimuli or events (Andrews & Robinson, 1991; Wilson & Gilbert, 2003). Some studies ask participants to predict their feelings to outcomes of genuine future events such as football games or presidential elections (Wilson et al., 2000; Scheibe, Mata, & Carstensen, 2011). Others use laboratory tests that entail verbal descriptions of emotional scenes and require participants to predict how they would feel subsequently when viewing the actual scene (Robinson & Clore, 2001; Hoerger, Chapman, Epstein, & Duberstein, 2012). In each case, AF ability is assessed by measuring the discrepancy between ratings of predicted feelings and ratings of feelings experienced at a later time, in order to determine the prediction-accuracy score, also referred to as absolute accuracy (Wilson & Gilbert, 2003; Mathieu & Gosling, 2012). Participants may be asked to predict how long or how frequently they will feel a given emotion or the degree to which an event or stimulus will affect their mood. Any distinct feature of predicted emotion can be used as criteria for AF accuracy.

The present study used ratings of emotional intensity, or predictions of how strongly participants will feel in response to a particular event. The tendency to overestimate the emotional impact, including the emotional intensity, of a future outcome is known as the “impact bias” (Wilson & Gilbert, 2003). For example, impact bias may be demonstrated by a student who overestimates how positive they expect to feel if they were to receive a high exam score. Some evidence suggests that emotional intensity is only moderately overestimated (Mathieu & Gosling, 2012) and other work suggests the bias may disappear depending on how emotional ratings are collected (Levine, Lench, Kaplan, & Safer, 2012). Still, others maintain that the impact bias is “alive and well” (Wilson & Gilbert, 2013). Regardless, there is substantial evidence for individual differences in AF performance (Dunn, Brackett, Ashton-James, Schneiderman, & Salovey, 2007, Hoerger et al., 2012; Hoerger, Chapman, & Duberstein, 2016) for which the mechanisms are not yet understood. To the extent that AWM plays a mechanistic role in AF, AWM may be an underlying source of inter-individual variability in AF. Therefore, the primary goal of the present investigation was to determine whether AWM is related to, and predictive of, individual differences in forecasting of emotional intensity.

Working memory is a limited capacity system that temporarily maintains information actively in mind (Baddeley, 1992, 2007, 2012) in support of goal-directed behavior, planning and problem-solving (Smith & Jonides, 1999). According to current models, working memory is composed of separable, domain-specific subsystems specialized for the short-term maintenance of different types of information (e.g., visual, verbal, spatial) (Repovš & Baddeley, 2006; Baddeley, 2012). Affective working memory is posited to be an additional domain-specific subsystem specialized for maintaining emotional feelings actively in mind over a brief interval (Mikels & Reuter-Lorenz, 2019; cf. LeDoux & Brown, 2017). Previous studies have measured AWM using an affect maintenance task in which participants view an emotional image, then assess and hold their emotional reaction (i.e., the intensity of their feelings) in mind over a delay. Participants then view another emotional image and determine if the second image evoked an emotional response with higher or lower intensity than the first picture in the pair (Mikels et al., 2005, 2008; Broome, Gard & Mikels, 2012). To determine if the ability to maintain an emotional feeling can be dissociated from working memory for affectively "neutral" information, previous investigations have used an analogous visual working memory task that measures non-affective maintenance abilities. In this task, participants assess and hold the *brightness intensity* of a neutral image in mind over a delay before deciding if a second image has higher or lower brightness intensity (Mikels et al., 2005, 2008; Broome et al., 2012). The brightness maintenance task is particularly well-suited for comparison with the affect maintenance task; both require the subjective assessment of pictorial stimuli (IAPS images; Lang, Bradley & Cuthbert, 1999) and a response decision based on assessing the relative intensity of subjective states evoked by two consecutive stimuli offset by a brief delay (Mikels et al., 2008). Furthermore, accuracy on both tasks can be scored using subjective intensity ratings. Following the maintenance phase of each task, participants rate the intensity (emotional or brightness) of each image used in the maintenance tasks. These ratings are then used to establish individualized accuracy scores for the respective maintenance task.

Several lines of research using variants of these tasks provide converging evidence that AWM is a separable subsystem that is at least partially dissociable from working memory for non-emotional, visual information (Mikels et al., 2005, 2008; see also Gard et al., 2011). Using selective interference methodology, Mikels and his colleagues (2008) found that a secondary emotion regulation task impaired performance on the affect maintenance task but not the brightness maintenance task. Conversely, a secondary cognitive task impaired performance on the brightness maintenance task but facilitated performance on the affect maintenance task, suggesting the dissociability of working memory for affective versus non-affective information (Mikels et al., 2008). Further evidence for this conclusion derives from findings that older adults performed comparably to younger adults on the affect maintenance task, but they were significantly impaired on the brightness maintenance task (Mikels et al., 2005). This pattern is consistent with the relative preservation of affective processing abilities in older adults despite cognitive decline that accompanies normal aging (Mather, 2016). Taken together, these studies suggest that AWM is dissociable from non-affective working memory, a property that was tested further in the present investigation.

Additionally, prior research points to a potential relationship between AF and emotional intelligence (EI)—a set of skills related to the assessment, management, regulation, and expression of emotion, and the use of feelings to succeed in everyday life (Dunn, Brackett, Ashton-James, Schneiderman, & Salovey, 2007; Hoerger et al., 2012; Salovey, Mayer, & Caruso, 2004). There are a variety of models (Petrides & Furnham, 2001; Salovey et al., 2004) and measures of the EI construct (see Hughes & Evans, 2018 for a recent review). Previous studies have identified relationships between AF and EI (albeit with some inconsistency) and these were generally hypothesized to be due to differences in basic emotion processing—such as the ability to identify, remember, and manage emotions (Dunn et al., 2007; Hoerger et al., 2102)—which could also relate to AWM. Thus, an ancillary goal of the present investigation was to reassess the relationships between trait and ability EI and AF, and to explore the potential relationships between AWM and EI. We expected that a better ability to maintain feeling states would be associated with superior EI and more successful AF.

In the first set of studies, 1a and 1b, we examined the relationship between AF and AWM, and between AF and visual working memory, respectively, in two independent samples of participants. We predicted there would be a positive relationship between AWM and AF (Study 1a), but no relationship between visual working memory and AF (Study 1b). Study 2 sought to replicate this selectivity within a new group of participants. Study 3 aimed to strengthen the evidence by replicating and extending the results using a pre-registered study. Evidence for these predictions would lend credence to the idea that AWM is a fundamental capacity that contributes to the accuracy of affective forecasts and provide additional evidence that AWM is dissociable from working memory for non-affective information.

Study 1a: Testing the relationship between AWM and AF Ability

Method

Participants

Seventy-nine undergraduate students participated in the study in exchange for course credit. One participant chose not to continue in the study after seeing the sample emotional images, three participants did not return for Session 2 (see Design and Procedures) and one participant failed to follow task instructions, resulting in the exclusion of these five participants. In addition, eight participants were excluded due to poor performance on the affect maintenance task (see below for details). Thus, the reported analyses were performed on data from 66 participants (63.5% female, mean age 18.83, 71.6% self-identified White), who self-reported as right-handed and native English-speakers. A power analysis based on data from Hoerger et al. (2012) Study 2 (N = 430), that used a similar AF task to measure the relationship between AF and a different variable of interest (i.e., emotional intelligence; see below), indicated we needed 46 participants to have 80% power for detecting a small-medium effect with the traditional alpha = .05 criterion of statistical significance (G*Power 3.1: Faul, Erfelder, Buchner & Lang, 2009). Our initial sample size exceeded the estimate to account for expected attrition and low performing individuals. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Design and Procedures

Participants completed two one-hour testing sessions, scheduled one week apart (see Figure 2-1). During the first session, participants performed 40 trials of the affect maintenance task, followed by Phase I of the AF task. Phase I of the AF task was the prediction phase, where participants were instructed to predict their future feelings based on a written description of a scene. During the second session, participants completed Phase II of the AF task, where participants were instructed to rate their experienced feelings after viewing an image of the scene. After Phase II of the AF task, participants completed another 40 trials of the affect maintenance task. This task sequence was intended to minimize the influence of the affect maintenance task on emotional ratings of the AF task between Phase I and Phase II. Finally, participants rated the emotional intensity of all images included in the maintenance task. All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

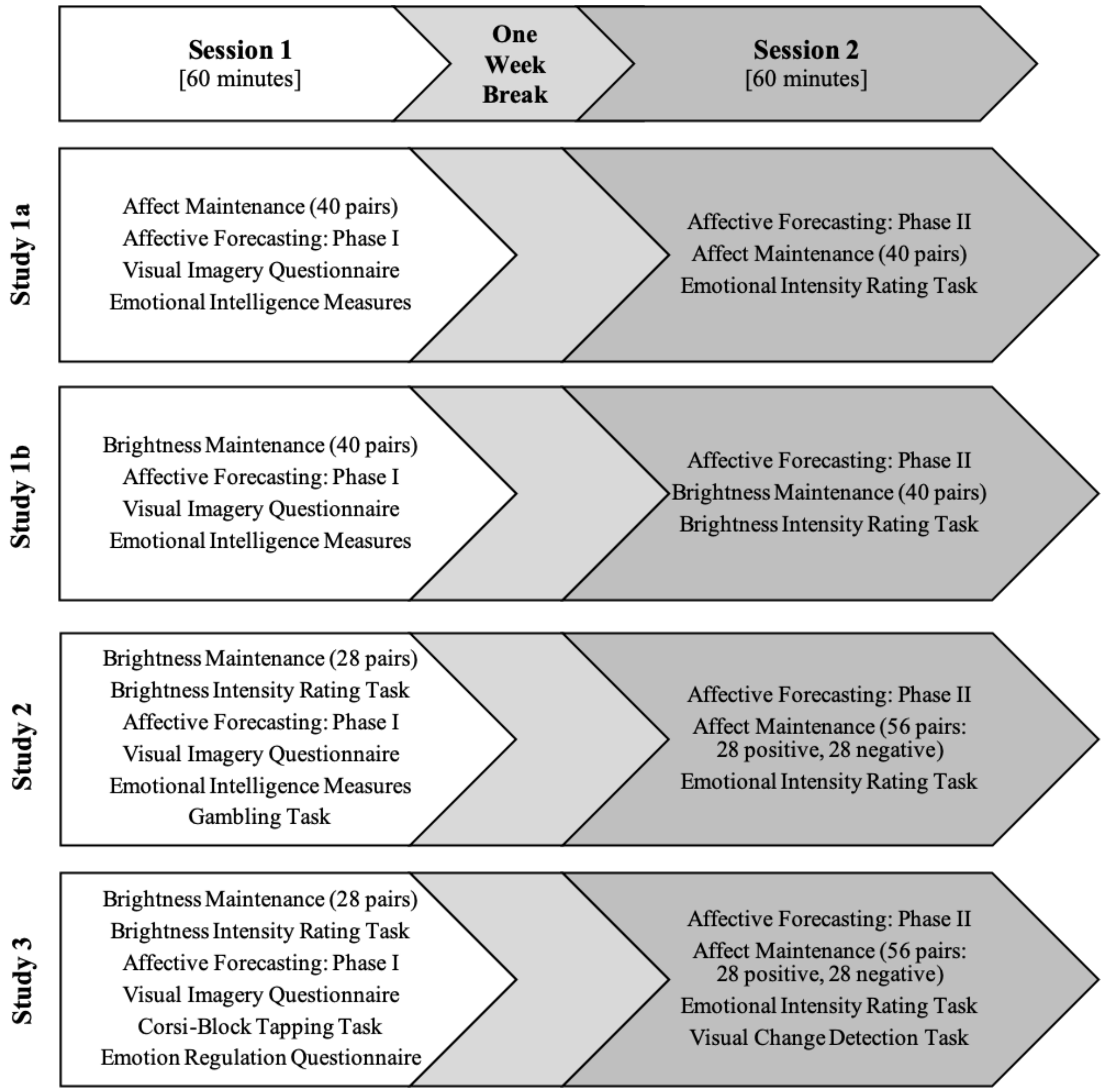

Figure 2-1. Protocols for Study 1a, Study 1b, Study 2, and Study 3.

Affect Maintenance Task. The affect maintenance task (Figure 2-2) was similar to that employed in previous research measuring AWM (Mikels et al., 2005, Mikels et al., 2008, Broome et al., 2012). For each trial, participants viewed one emotional picture (5s) and were instructed to maintain the feelings elicited by this image. A retention interval ensued (3s) before participants viewed a second emotional picture (5s). Next, a green cross appeared, which prompted participants to report whether the second image had higher or lower emotional intensity than the first one. Emotional intensity was described to the participants as the strength or magnitude of their emotional reaction to each image, regardless of the picture's content. Participants responded to each trial by pressing either a key labeled "H" for higher or "L" for lower.

There were 80 trials total and each trial consisted of a pair of images selected to have matching valence (40 positive and 40 negative trials). We used the stimuli selected by Broome et al. (2012), consisting of 80 matched-valence pairs created from a set of 160 images (80 positive, 80 negative). Images depicted pleasant (positive valence) or unpleasant (negative valence) scenes selected from the IAPS (Lang et al., 1999) and supplemented from an in-house database to include 17 additional high arousal images (available upon request). Intensity ratings from all images were originally obtained from two independent samples (N_1 = 40 & N_2 = 40; Mikels et al., 2008) and were recorded using a 7-point scale that ranged from low (1) to high (7) emotional intensity, or degree of emotional reaction to each image. The difference in emotional intensity ratings between the two images in these pairs varied from 0.23 to 1.75 (M = .99, SD = .47). As in previous work with this image set, this difference, referred to as intensity distance, was used to categorize the stimulus set into near (intensity distance .88 or below) and far (intensity distance 1.03 or greater) pairs, resulting in 38 near pairs and 42 far pairs. Additionally, for each intensity distance subset, the second picture had higher intensity than the first for exactly half of the pairs, and vice versa for the other half. Following Broome et al., (2012), the 80 pairs were divided into two blocks (Image Sets A and B) such that they were equated for valence, intensity distance (near or far) and intensity order (second picture higher or lower), and their presentation at each of the two sessions was counterbalanced across participants. These restrictions were to ensure there would be no differences in performance due to the image sets, which was later confirmed by our participants, $t(65) = .46$, $p = .63$. Trials within each block were presented in a randomized order for each participant.

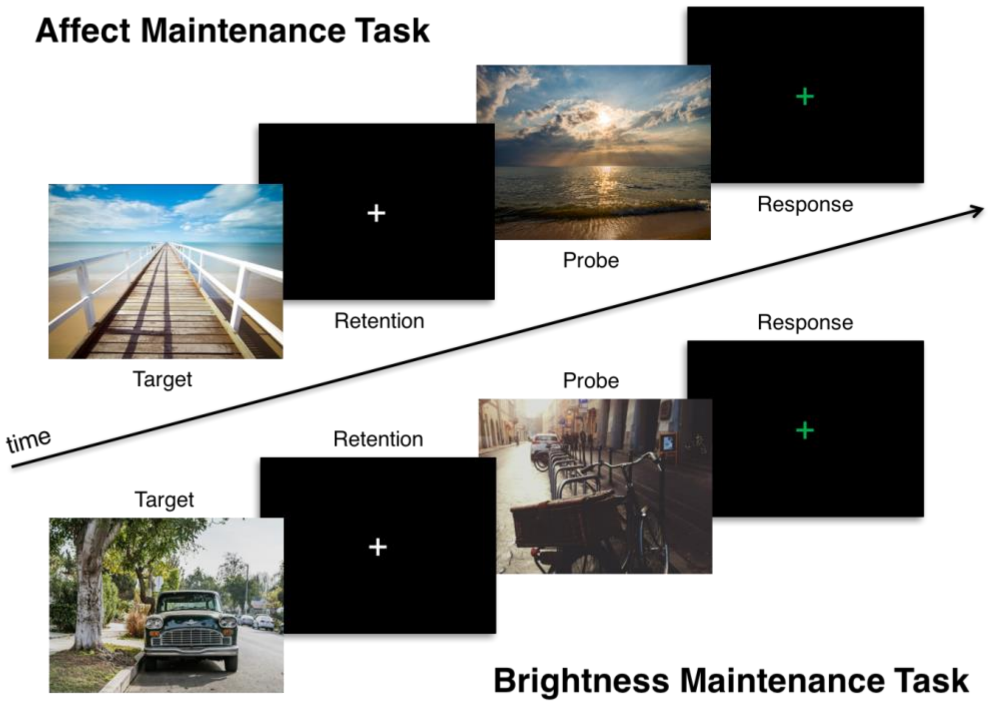

Figure 2-2. Schematic of maintenance tasks used in Studies 1–3. Participants hold the emotional or brightness intensity of one image in mind over a delay to determine if a subsequent image has higher or lower intensity. Image adapted from "Affective Working Memory: An Integrative Psychological Construct" by J. A. Mikels and P. A. Reuter-Lorenz, 2019, Perspectives in Psychological Science, 14(4), p. 8. Copyright 2019 by SAGE Publications.

Affective Forecasting. During the first session of the AF task (i.e., Phase I; see Figure 2-3), participants read a description of a scene and were asked to imagine it. Participants then predicted how they would feel if they were to view the actual image and rated this feeling on a visual analog scale (actual resolution: 21 points) that ranged from endpoint anchors labelled "very unpleasant" to "very pleasant". Participants read ten such descriptions and rated their predicted feelings for each by clicking the computer mouse anywhere along the scale. During

Session 2, one week later (i.e., Phase II; see Figure 2-3), participants viewed the images associated with each description and rated how each one made them feel using the same scale. Five negative (mean pleasure rating = 2.35, *SD* = 0.68) and five positive (mean pleasure rating = 7.42, *SD* = 0.69) images from the International Affective Picture System (IAPS, Lang et al., 1999) were used, based on those used in Robinson & Clore (2001). Five of the images were the same as in Robinson and Clore (2001); to avoid repeating images already employed in the affect maintenance task, the other five images were replaced with alternative IAPS images matched as closely as possible in content to those used by Robinson and Clore (2001). Accordingly, five descriptions were taken from Robinson and Clore (2001) and the remaining were constructed to match the new replacement images (see Table 2-3 in Appendix 2A for image descriptions used in the current study). Descriptions and images were displayed on a desktop computer screen, and the order of stimuli was randomized for each participant independently for Phase I and Phase II.

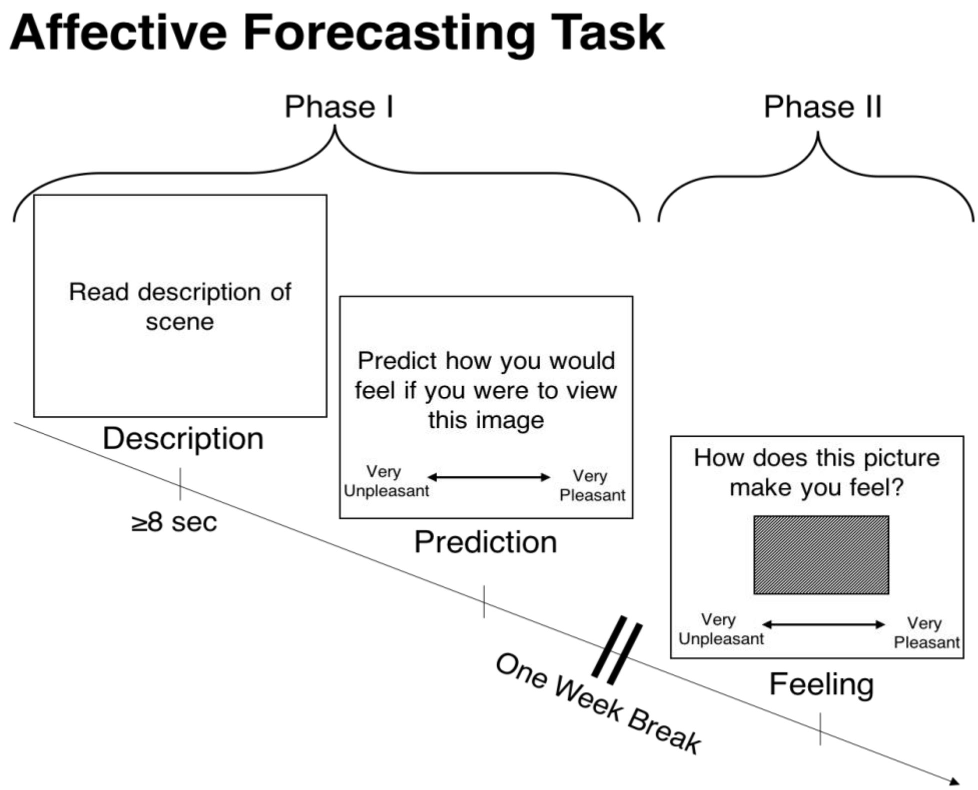

Figure 2-3. Schematic of the affective forecasting task used in Studies 1–3. During Phase I, participants read brief affectively laden descriptions of scenes and predict how they would feel if they were to view each image described. After a one-week break, during Phase II, participants view each image and rate their experienced feelings.

Emotional Intensity Ratings. Accuracy on the affect maintenance task was scored on an individualized basis using each participant's subjective ratings of emotional intensity for each image (see Affective Forecasting and Maintenance Task Scoring for details). In order to obtain subjective ratings of emotional intensity, participants viewed all images from the affect maintenance task again individually, at the end of the second session. The rating task was performed at the end of session two in order to minimize the possibility that the process of rating the images or that memory of prior ratings would contaminate relative intensity judgments and influence performance on the affect maintenance task. Images were randomly assigned to one of four blocks and trials within each block were presented in a randomized order. Participants were reminded that emotional intensity referred to the strength or magnitude of their emotional reaction. Then, using a visual analog scale that ranged from "not intense" to "extremely intense," displayed below the image, participants rated the intensity of their feelings for each image. Responses were registered by a mouse click anywhere on the scale corresponding to the intensity of their feelings. The scale had a resolution of 21 units and ratings were used to calculate affect maintenance accuracy scores, as described below (see Affect Maintenance Accuracy Scores).

Additional Measures

We assessed the participant's visual imagery ability to control for its potential influence on the AF task. Due to significant relationships between emotional intelligence (EI) and AF found in previous studies, we also included trait and ability measures of emotional intelligence to see if the results would replicate here (Dunn et al., 2007; Hoerger et al., 2012). These additional measures were administered at the end of the first testing session (see Figure 2-1 for the complete order of tasks). We also assessed the participants' emotional state before and after each session to ensure they were not negatively influenced by the stimuli viewed in the study. These measures are briefly described, in turn, below.

Visual Imagery. Visual imagery was assessed with the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). The self-report measure asks participants to answer four questions about each of four imagined scenes, and to rate vividness on a 5-point scale of vividness. A response of '1' represents "perfectly clear and as vivid as normal vision" and '5' represents "no image at all, you only *know* that you are thinking of the object." Scores for all 16 items were summed to create a composite VVIQ score where higher scores indicate poorer imaging capabilities (scores range from 16 to 80).

Emotional Intelligence. Based on previous findings revealing relationships between AF and EI, we employed the Situational Test of Emotional Understanding (STEU; α = .77; range: 0 – 1, MacCann & Roberts, 2008) to assess ability EI and the Assessing Emotions Scale (AES; α = .87 – .90; scores range from 33 to 165, Schutte et al., 1998), also known as the Schutte Emotional Intelligence Scale (Schutte, Makiyff, & Bhullar, 2009), to assess trait EI.

Emotional State. To assess potential changes in emotional state due to participating in the study, the "state" version of the Positive and Negative Affect Scale (PANAS; positive items α = .89, negative items α = .85; Watson, Clark and Tellegen, 1988) was administered at the beginning and end of each testing session. The scale consists of twenty words that describe feelings and emotions. For each word, participants are asked to rate the extent they feel that way "right now, that is, at the present moment" from 1 (not at all) to 5 (extremely). This measure was used only to ensure that the subjects' emotional states were not negatively impacted by the stimuli shown in our experiment and will not be discussed further.

Affective Forecasting and Maintenance Task Scoring

Affect Maintenance Accuracy Scores. Accuracy scoring was individualized based on each participant's image ratings from the emotional intensity rating task. These subjective intensity ratings were used to infer which image, when appearing with its pair, would be judged by that individual as having higher intensity. A response on the affect maintenance task was scored correct when it agreed with the participant's relative intensity ratings for the members of each pair. Trials with images receiving equal intensity ratings were excluded from the calculation of that participant's accuracy scores. On average, approximately six of each participant's 80 total trials, or 7.8% (*SD* = 3.02%) of all trials, were excluded for this reason. For each participant, accuracy scores were calculated as the sum of the affect maintenance responses that matched the intensity-based comparisons and divided by the number of trials included in the sum. Separate accuracy scores were calculated for each session and then averaged together to calculate a composite affect maintenance accuracy score. As mentioned above, eight participants were excluded from all subsequent analyses because they performed more than two standard deviations below the sample mean during either session. This level of performance was approximately 50% accuracy, suggesting participants were merely guessing, or otherwise not following instructions.

AF Scores. AF prediction-accuracy scores were calculated based on the difference between the predicted rating in response to the picture description and the feeling rating in response to viewing the image for each participant. The absolute value of this difference was subtracted from the largest possible deviation score (i.e., 10), thus providing an AF score between zero (least accurate prediction) and 10 (most accurate prediction). For each participant, we averaged the scores across the ten images to calculate a composite AF score (see also Dunn et al., 2007; Hoerger et al., 2012). Because we hypothesized that AWM would contribute to AF accuracy per se, we focused on magnitude, rather than the direction (i.e., overestimation vs. underestimation) of AF errors (see Appendix 2B for exploratory analyses on directionality).

Results

Descriptive statistics for affective forecasting (AF), affective maintenance, and vividness of visual imagery (VVIQ) are presented in Table 2-1. As can be seen from this table, in this and all subsequent studies, participants performed the affective maintenance task reasonably well, and were comfortably below ceiling. Likewise, AF performance was on average reasonably accurate but falling below a perfect 10 by approximately 2.5 points.

Table 2-1. Mean (SD) Performance on Key Measures Across Studies 1, 2, & 3

Measures	Study 1a N = 66	Study 1b N = 68	Study 2 N = 96	Study 3 N = 85
Affective Forecasting	7.48 (0.85)	7.35 (0.90)	7.61 (0.89)	7.70 (0.78)
Affect Maintenance	0.77 (0.06)	NA	0.81 (0.09)	0.81 (0.08)
Brightness Maintenance	NA	0.78 (0.08)	0.85 (0.08)	0.80 (0.08)
VVIQ	38.6 (9.08)	39.4 (9.48)	37.5 (9.85)	38.5 (10.8)

Range of Scores. Affective Forecasting: 0–10, Affect Maintenance: 0–1, Brightness Maintenance: 0–1, and VVIQ: 16–80, where a score approaching 16 reflects the greatest reported imaging ability and a score approaching 80 reflects the poorest abilities.

A multiple linear regression tested whether affective maintenance accuracy and VVIQ scores predicted AF performance. Results indicated this model was significant, ($F(2,63) = 5.47$, $p = .006$) with an $R^2 = .15$. As predicted, affect maintenance accuracy contributed significantly to the model, $\beta = 0.32$, $p = .008$ (see Figure 2-4). Additionally, VVIQ scores also contributed significantly ($\beta = –0.24$, $p = .043$) such that better visual imaging abilities predicted higher AF accuracy.

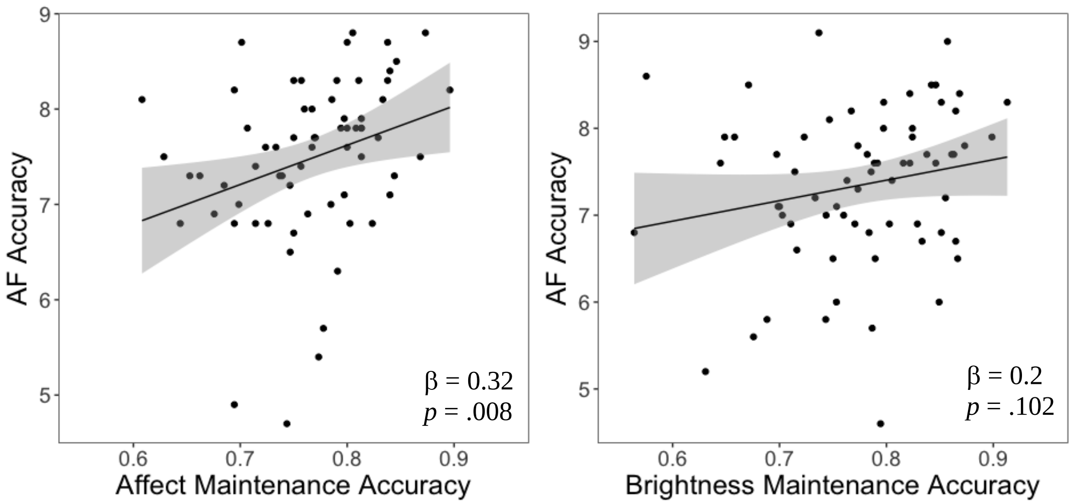

Figure 2-4. Scatterplots showing forecasting accuracy plotted as a function of maintenance task performance in Study 1. The graphs (with best fitting regression line and 95% confidence interval shaded region) show that AF accuracy is significantly predicted by affect maintenance accuracy (left; Study 1a) but not brightness maintenance accuracy (right; Study 1b) using independent samples.

Ancillary Analyses

In an effort to replicate correlations found in previous research, the relationship between AF and EI was assessed using Pearson's r correlation coefficient (see Table 2-2 for EI scores). Neither the relationship between AF and trait EI ($r(62) = .06$, $p = .62$) nor between AF and ability EI were significant, $r(64) = –.04$, $p = .75$. Affect maintenance accuracy was also unrelated to either measure of EI (trait: $r(62) = –.13$, $p = .32$; ability: $r(64) = –.07$, $p = .55$).

Table 2-2. Mean (SD) Performance on the Emotional Intelligence Measures Across Studies 1 & 2

Measures	Study 1a *N* = 66	Study 1b *N* = 68	Study 2 *N* = 96
EI – Ability	.68 (.08)	.68 (.1)	.68 (.07)
EI – Trait	120.8 (8.3)	122.0 (12.3)	123.0 (11.9)

Range of Scores. EI–Ability: 0–1, EI–Trait: 33–165

Discussion

Results from the first study, while correlational, provide initial support for the hypothesis that AWM plays a role in AF. The positive relationship observed between emotion maintenance ability and AF motivates the question of specificity: Is the association between AWM and AF due to the short-term maintenance of emotional information per se, or does it reflect a more general role for working memory? Therefore, Study 1b examined the relationship between AF and visual working memory using the brightness maintenance task originally designed as the non-affective analogue of the affect maintenance task. It requires participants to maintain in mind the brightness intensity of an emotionally neutral image over a delay, to compare with the brightness of a subsequent neutral image (Mikels et al., 2005, Mikels et al., 2008; Broome et al., 2012). In contrast to affect maintenance, we predicted that the ability to maintain representations of brightness intensity would have *no* relationship with AF ability. Study 1a also found a correlation between AF and visual imagery as measured by the VVIQ. This measure is included in Study 1b and subsequent studies to assess the reliability of this relationship.

Study 1b: Testing the relationship between Visual Working Memory and AF Ability

Method

Participants

A new sample of seventy-eight undergraduate students participated in Study 1b. Sixty-nine participated in the study in exchange for course credit, while nine participated for monetary compensation in order to continue data collection during a university break. Three participants did not return to complete Session 2 and were therefore excluded. In addition, data from five participants were excluded due to poor performance on the brightness maintenance task (see below for details), and one participant was excluded for being under the age of 18 at the time of testing, in accordance with our inclusion criteria. Thus, the present analyses were performed on data from 68 participants (66.2% female, mean age 18.86, 67.6% self-identified White), who self-reported as right-handed and native English-speakers. The sample size was selected to parallel Study 1a, over-recruited to account for attrition and low-performing participants. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Design and Procedures

Participants completed two one-hour testing sessions, scheduled one week apart (see Figure 2-1). The experimental design and tasks were the same as Study 1a, except for the use of the brightness maintenance task and corresponding brightness rating task. All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013).

Brightness Maintenance Task. In the brightness maintenance task, participants viewed one neutral image (5s) and were instructed to maintain the brightness intensity actively in mind. A retention interval ensued (3s) before participants viewed a second neutral image (5s). Next, a green cross appeared, which prompted participants to report whether the second image had higher or lower brightness intensity than the first one. Brightness intensity was described to participants as the magnitude of overall light or illumination in the image, regardless of the picture's content. Participants responded to each trial by pressing either a key labeled "H" for higher or "L" for lower (Figure 2-2). There were 80 trials total where each trial consisted of one pair of two neutral images. We used the same 80 pairs as in Broome et al. (2012). Images were neutral scenes selected from the IAPS (Lang et al., 1999) to ensure that performance would be independent from emotional processes. Intensity ratings from images were obtained previously (N = 40; Pilot Study C, Mikels et al., 2008) and recorded using a 7-point scale that ranged from low (1) to high (7) brightness intensity, or overall light or illumination of each image. The difference in brightness intensity between images in these pairs, or intensity distance, varied from 0.10 to 2.98 (M = 1.28, SD = .80) and consonant with previous work using this image set, pairs were divided into two groups: near (intensity distance 1.13 or below) and far (intensity distance 1.15 or greater), resulting in 40 near pairs and 40 far pairs. Additionally, for each intensity distance subset, the second picture would have higher intensity than the first for exactly half of the pairs. Furthermore, to establish whether image order affected performance, half of the participants viewed one intensity order and half viewed the same pairs with the image order reversed. This manipulation had no effect on performance and will not be discussed further, $t(66) = 0.59$, $p = .555$. Each order variant of the 80 pairs was divided into sets (Set A and Set B) with equal number of trials for intensity distance (near and far) and intensity order (second image higher or lower intensity) and their presentation at each of the two sessions was counterbalanced across participants. This was intended to minimize performance differences due to the image sets, and proved to be effective, $t(67) = -1.18$, $p = .244$. Within these restrictions, pair order was randomized individually for each participant.

Brightness Intensity Ratings. At the end of the second session, participants viewed all images from the brightness maintenance task again and rated the “intensity” or magnitude of overall light or illumination they perceived for each one. Ratings were collected after the maintenance task to avoid the potential influence of prior ratings on brightness maintenance performance. Each picture was displayed individually above a visual analog scale that ranged from “not intense” to “extremely intense”. Participants responded by clicking anywhere on the scale below the picture to indicate the intensity of their perceived brightness for each image. The scale had an actual resolution of 21 units and ratings were used to calculate the brightness maintenance accuracy scores (see Brightness Maintenance Accuracy Scores below).

Maintenance Task Scoring

Brightness Maintenance Accuracy Scores. Similar to the calculation of affect maintenance accuracy, brightness maintenance accuracy scoring was individualized using each participant's ratings from the brightness intensity-rating task. These subjective ratings were used to infer which image, when appearing with its pair, would be judged by that individual as having higher intensity. A response in the brightness maintenance task was considered correct when it agreed with the participant's relative intensity ratings for the members of each pair. Trials with images receiving equal intensity ratings were excluded from the calculation of that participant's accuracy scores (M = 7.4% (SD = 4.0%) of all responses). For each participant, accuracy scores were calculated as the sum of the brightness maintenance responses that matched the intensity-based comparisons and divided by the number of included responses. Separate accuracy scores were calculated for each session and then averaged together to calculate a composite brightness maintenance accuracy score. As mentioned above, six participants were excluded from analyses because they performed more than two standard deviations below the mean during either session. These participants performed at approximately 50% accuracy, suggesting they were guessing or otherwise not following instructions.

Results

Descriptive statistics for affective forecasting (AF), brightness maintenance task performance, and vividness of visual imagery (VVIQ) are presented in Table 2-1. Performance on the brightness maintenance task was reasonable, below ceiling, and well matched with accuracy on the affective maintenance task in the prior study. AF performance was similar to that in Study 1a.

A multiple linear regression was computed to determine if brightness maintenance accuracy and VVIQ scores were significant predictors of AF performance. This model was not significant, ($F(2, 65) = 1.34$, $p = .268$) with an $R^2 = .04$, and neither brightness maintenance performance ($\beta = 0.2$, $p = .108$; see Figure 2-4) nor visual imagery ability ($\beta = -0.007$, $p = .953$) contributed significantly to the model.

Ancillary Analyses

As in Study 1a, we tested the relationships between AF and EI and found that neither the relationship between AF and trait EI ($r(66) = .15$, $p = .21$) nor between AF and ability EI ($r(66) = .14$, $p = .26$) was significant. Brightness maintenance accuracy was not related to trait EI ($r(66) = -.09$, $p = .48$) but brightness maintenance accuracy and ability EI were related, $r(66) = .30$, $p = .01$.

Discussion

As predicted, the results from Study 1b provide no indication that brightness maintenance ability is related to AF. Note also that the VVIQ was not associated with AF in this sample. Separately then, the results of Studies 1a and 1b demonstrate that AF ability is selectively related to affect maintenance. However, because these relationships were examined using two different samples, it was not possible to assess the contribution of one maintenance ability while adjusting for the effect of the other. Therefore, the hypothesis that predicting future feelings is specifically related to affective working memory ability was further tested in a second study using both maintenance tasks in a within-subjects design. Study 2 aimed to replicate the specific relationship observed in Study 1 between AF and affect maintenance and allowed us to compare the extent to which performance on each maintenance task predicted AF ability. We expected that only affect maintenance would be a significant predictor of AF ability, whereas brightness maintenance would not.

Study 2: The Contributions of Affective and Visual Working Memory to AF Ability Compared

Method

Participants

A new sample of one-hundred and ten undergraduate students participated in Study 2 for course credit. One participant chose not to continue in the study after seeing the sample emotional images, two participants did not return for Session 2, and four participants failed to follow instructions, resulting in the exclusion of these seven participants. Due to experimenter error, another participant performed an incorrect version of the task and their data was excluded. Additional exclusions included five participants who did not meet performance inclusion criteria for the maintenance tasks as described in Studies 1a and 1b (see below for details) and one clearly outlying participant who performed more than 3.5 standard deviations below the mean on the AF task (Note: When this 3.5 SD criterion for the AF was retroactively applied to Studies 1a and 1b, no outliers were identified). Thus, the present analyses were performed on data from 96 participants (63.5% female, mean age 18.8, 67% self-identified White), who were self-reported right-handers and native English-speakers. A power analysis based on predictor correlations from Studies 1a and 1b indicated 81 participants would yield 80% power for detecting a small-medium effect at the traditional alpha = .05 criterion of statistical significance (G*Power 3.1: Faul, Erfelder, Buchner & Lang, 2009). We oversampled to account for attrition and low performing individuals. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Design and Procedures

All participants performed tasks used in Studies 1a and 1b to measure affective maintenance, brightness maintenance, and AF so that the relationships among these constructs could be examined. Participants completed two one-hour testing sessions, scheduled one week apart. The task order used in the prior studies was adjusted in Study 2 to ensure that all assessments could be completed within the one-hour per session design (see Figure 2-1). All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

Brightness Maintenance Task. In order to maximize efficiency of the brightness maintenance task, we first identified and selected pairs that were most discriminative of brightness maintenance ability. While all pairs previously included in the brightness maintenance task were originally rated as low arousal with intermediate valence (Pilot Study C; Mikels et al., 2008), an additional assessment made by two independent raters excluded appetitive stimuli (e.g. food or drink) and potentially non-neutral content (e.g. lava or lightening). From the remaining 46 pairs, we used an IRT analysis of the data from Study 1b to identify the 28 pairs with the highest discrimination scores, which were presumably the most sensitive to maintenance ability (See Appendix 2C for full method on IRT analyses). Apart from the number of trials, the brightness maintenance task was the same as Study 1b.

Brightness Intensity Ratings. Immediately after the maintenance task in the first session, participants viewed all images from this task again and rated their brightness “intensity” or overall magnitude of perceived illumination for each image individually. Due to a programming error, one pair presented during the brightness maintenance task failed to be presented during the rating task, preventing us from scoring this trial. Thus, this pair was dropped when calculating the overall accuracy scores and only performance from the completed 27 pairs was used.

Affective Forecasting. As in the previous studies, during the first session, participants read descriptions of emotional images and were asked to predict how they would feel if they were to view each image (i.e., Phase I). Then, during the second session a week later, participants saw the images and rated how they actually felt using the same scale (i.e., Phase II). The descriptions, images, protocol, and scoring procedure were identical to those used in Studies 1a and 1b.

Affect Maintenance Task. In an effort to maximize efficiency of the affect maintenance task, we identified and selected the pairs that were most diagnostic of AWM ability. Using an Item-Response Theory (IRT) analysis, we selected 56 pairs from those used in Study 1a (see Appendix 2C for details). We chose the 28 most discriminative positive pairs and the 28 most discriminative negative pairs using the data from Study 1a. This yielded the subset of pairs that were the most diagnostic of AWM ability for each valence. All pairs for the affect maintenance task were viewed during the second session. Parameters of the task, except for the number of pairs, remained identical to Study 1a.

Emotional Intensity Ratings. Immediately after the affect maintenance task in Session 2, participants viewed each image from the affect maintenance task again individually and rated the “intensity” or overall magnitude of emotional reaction for each. The design of this task was the same as Study 1a.

Additional Measures

Similar to Studies 1a and 1b, participants completed the VVIQ (Marks, 1973), STEU (MacCann & Roberts, 2008), and AES (Schutte et al., 1998) during the first session. Participants also performed the PANAS (Watson, Clark and Tellegen, 1988) at the beginning and end of both sessions to ensure their emotional states were not negatively influenced by the images viewed in the study. For exploratory reasons we included a gambling task adapted from De Martino et al. (2006; see Mikels & Reed, 2009), which was administered at the very end of Session 1 and will not be discussed.

Maintenance Task Scoring

Affect and Brightness Maintenance Accuracy Scores. The accuracy for each task was determined using the individualized scoring procedure described for Studies 1a and 1b. As mentioned above, five participants were excluded from all analyses because they performed more than two standard deviations below the mean on at least one of the maintenance tasks (i.e., two participants met this criterion for the affect maintenance task and three for the brightness maintenance task).

Results

Descriptive statistics for affective forecasting (AF), affect maintenance and brightness maintenance task performance, and vividness of visual imagery (VVIQ) are presented in Table 2-1. As the means indicate, brightness maintenance accuracy ($M = .85$, $SD = .08$) was slightly but significantly better than affect maintenance accuracy ($M = .81$, $SD = .09$; $p < .05$), a difference that is likely due to pair selection, as we address further below. Because linear regression accounts for predictor variables with different means, we proceeded with our planned tests. Additionally, AF performance was consistent with the previous results.

A multiple linear regression was computed with AF accuracy as the outcome variable and affect maintenance accuracy, brightness maintenance accuracy, and VVIQ scores as predictors. Results indicated that this model was significant ($F(3,92) = 5.37$, $p = .002$) with an $R^2 = .15$. As predicted, affect maintenance performance contributed significantly to the model ($\beta = 0.33$, $p = .001$) whereas brightness maintenance performance did not ($\beta = 0.16$, $p = .103$; see Figure 2-5). VVIQ scores did not contribute to the model, $\beta = 0.01$, $p = .89$. In order to examine the relationships between performance on each maintenance task and AF while holding the other constant, we ran partial correlations. The partial correlation between AF and affect maintenance controlling for brightness maintenance was significant ($\rho = .34$, $p <. 001$), whereas the partial correlation between AF and brightness maintenance controlling for affect maintenance was not significant ($\rho = .17$, $p = .101$). Additionally, no relationship between performance on the brightness maintenance and affect maintenance tasks was found, $r(94) = .12$, $p = .252$.

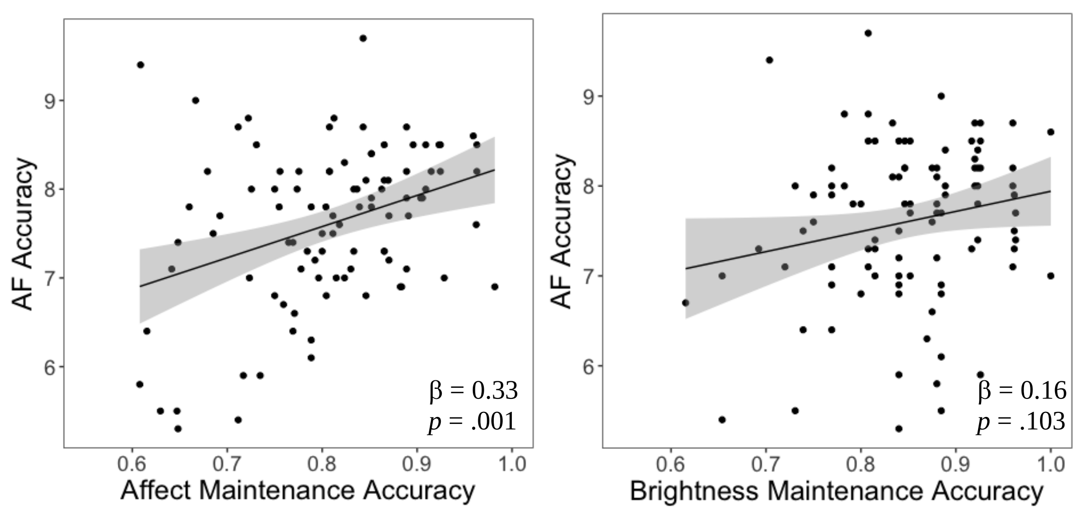

Figure 2-5. Scatterplots showing forecasting accuracy plotted as a function of maintenance task performance in Study 2. The graphs (with best-fitting regression line and shaded region indicating 95% confidence interval) show that Affect Maintenance Accuracy (left) but not Brightness Maintenance Accuracy (right) predicts AF Ability using a within-subject design (Study 2).

Ancillary Analyses

Using Pearson's r correlation coefficients, the relationships between AF and EI were assessed and no significant relationships between AF and either type of EI (trait: $r(93) = .09$, $p = .383$; ability: $r(94) = .18$, $p = .076$) were found (see Table 2-2 for EI scores). Affect maintenance accuracy was not related to trait EI ($r(93) = –.07$, $p = .521$) but there was a significant relationship between affect maintenance accuracy and ability EI, $r(94) = .31$, p = .002. Brightness maintenance accuracy performance was not related to trait EI ($r(93) = –.08$, $p = .452$) nor ability EI, $r(94) = .02$, $p = .857$.

Studies 1 & 2 Combined Results from the Additional Measures

To determine whether previously reported findings of a positive association between EI and AF replicated, we included measures of trait and ability EI in all current studies. Although no significant relationships were found between EI and AF in each study separately ($.04 \leq r$s $\leq .18$; all ps $\geq .08$), we combined the data sets to create a larger sample ($N = 230$ participants) and repeated the correlations with increased power. AF was not significantly related to either type of EI (trait: $r(225) = .10$, $p = .14$; ability: $r(228) = .11$, $p = .11$). Thus, no overall relationships were found between AF and either type of EI. Additionally, there was no significant overall relationship between VVIQ and AF such that self-reported visual imagery ability was not related to forecasting accuracy, $r(228) = –.03$, $p =.692$.

Discussion

The results from Study 2 replicate and extend the findings from Studies 1a and 1b by demonstrating that emotion maintenance ability predicts AF accuracy whereas brightness maintenance does not. Further, the analyses of the combined data from the three participant samples collected thus far provide no evidence that either measure of EI or visual imagery are related to AF.

Study 3 aimed to build on these results and take the effects a step further, replicating and extending the selective association between affect maintenance and AF using methods and analyses that have been pre-registered. This next study also included two additional, more widely used measures of visual working memory. Due to their non-affective nature, we expected them to be unrelated to AF, which would provide further evidence for specificity and strengthen the argument that working memory for emotion plays a unique role in AF. Accordingly, we hypothesized that affect maintenance performance would be a significant predictor of AF ability whereas performance on the brightness maintenance and other cognitive working memory tasks would not be.

Because of the difference in accuracy between affect and brightness maintenance in Study 2, in Study 3 we made several stimulus substitutions to better equate accuracy on two maintenance tasks. Of note, this unexpected accuracy difference in Study 2 is unlikely to have caused the selective association we observed between affect maintenance and AF, because linear regression accounts for predictor variables with different means.

Additionally, because the relationships between AF and EI were found to be unreliable in the prior studies, we opted not to include EI measures in Study 3. Instead, for exploratory purposes, we included a measure of emotion regulation (ER; Gross, 2013). Participants self-reported the tendency to regulate their emotions using (a) cognitive reappraisal, which refers to changing how one appraises internal or external situations in an effort to modify the emotional significance, and (b) expressive suppression, which refers to the inhibition of emotional expression (Gross, 2013). As previously described, Mikels et al. (2008) found that performance was impaired on the affect maintenance task while participants were performing a concurrent ER (cognitive reappraisal) task. This interference may be indicative of a shared underlying process. Therefore, we predicted a positive relationship between AWM and ER, specifically cognitive reappraisal.

Study 3: Replicating and Extending the Selective AWM-AF Association

Method

Participants

A new sample of ninety-six young adults participated in exchange for course credit or payment. Five participants failed to follow instructions or complete the study and six were excluded due to poor performance on the maintenance tasks (see below for details). Thus, the reported analyses were performed on data from eighty-five participants (72.6% female, mean age 18.7 (*SD* = 0.90), 70.0% white), who were self-reported right-handers and native English speakers. A power analysis based on Study 2 indicated that, in order to detect a medium-large effect with five predictors and 80% power at the traditional alpha = .05 criterion of statistical significance, we needed at least 79 participants (G*Power 3.1). We over-recruited for anticipated attrition and low-performing individuals. All experimental procedures were approved by the University of Michigan Institutional Review Board. The methods and analyses were pre-registered and can be found at http://aspredicted.org/blind.php?x=q6x38e (note: revised version available here http://aspredicted.org/blind.php?x=hp6zi4).

Design and Procedures

Similar to Study 2, all participants completed two one-hour testing sessions scheduled one week apart that included the affect maintenance, brightness maintenance and AF tasks (see Figure 2-1). Additionally, participants performed two visual working memory tasks, split between weeks to minimize mental fatigue (Session 1: Corsi Block-Tapping; Session 2: Visual Change Detection). All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

Maintenance Tasks. In Study 3 we made several substitutions to the stimulus set in an effort to better equate overall performance on the two maintenance tasks. Our prior approach was to select the top-most discriminative pairs for the affect and brightness tasks respectively (see Appendix 2C). Consequently, the brightness pairs were inadvertently more discriminative than the affect pairs. For Study 3, we selected again the most discriminative pairs for each domain, with the constraint that the average discrimination index across the pairs for each task would be statistically equivalent. This led to a change in 21 brightness pairs and 11 affect pairs. In all respects, the task parameters remained identical to Study 2.

Rating Tasks. Participants performed the corresponding rating task after each maintenance task. For emotional intensity ratings, participants viewed all images from the affect maintenance task again individually and rated the “intensity”, or overall magnitude of emotional reaction for each. For the brightness ratings, participants viewed all images from the brightness maintenance task again individually and rated the perceived “intensity”, or overall magnitude of light or illumination of the image. The task design was identical to Studies 1 and 2.

Affective Forecasting. The AF task was identical to the one previously used in Studies 1a, 1b, and 2.

Corsi Block Tapping Task. To assess visuospatial working memory, we employed a computerized version of forward and backward Corsi-Block Tapping Task (WMS-R; Wechsler, 2000). A subset of nine white blocks turn red sequentially on the computer screen in a particular sequence. In the task, participants then repeat the sequence by clicking on the squares in the same (forward) or reverse (backward) order. Set size (i.e., sequence length) increases from three to nine squares with three trials for each set size. The task is discontinued when the participant answers all three trials of a given set size incorrectly. Scores were calculated as the product of the largest set size attempted and the total number of correctly reproduced sequences, calculated separately for backward and forward versions of the task. These two scores were then standardized and averaged to obtain one composite Corsi score.

Visual Change Detection. A computerized delayed match-to-sample, visual change detection task (Luck & Vogel, 1997) served as an additional measure of visual working memory. In this task, participants view a sample array of two to ten variously colored squares (500ms). The squares then change to reveal variants of a horizontal stripe pattern that includes equal parts of the six possible colors the squares can be (500ms). A test probe then appears and participants must indicate whether the square in that position is the same color it was in the initial sample array. Cowan's capacity score (K) is calculated using the formula (hit rate + correct rejection rate – 1, for each set size) x N, where N is set size (Cowan, 2001; Rouder, Morey, Morey, & Cowan 2011). To avoid problematic averaging as advised in Rouder et al. (2011), we excluded sub-span capacities (i.e., N = 2) and negative K scores were changed to zero before averaging across set sizes to create a composite K score for each participant.

Additional Measures

Similar to Studies 1 and 2, participants completed the VVIQ (Marks, 1973). The Positive Affect and Negative Affect Scale (PANAS) was again administered before and after each session to ensure mood was not negatively affected by the study and was not analyzed further (Watson, Clark, & Tellegen, 1988). Additionally, at the end of Session 1, participants completed a self-report measure of emotion regulation strategies, described next.

Emotion Regulation Questionnaire (ERQ). To explore potential relationships between emotion regulation, AWM, and AF, participants completed the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). The ERQ measures the tendency of participants to regulate their emotions using Cognitive Reappraisal and Expressive Suppression, two ER strategies. Participants read ten statements regarding how individuals regulate and manage their emotions and rated the extent to which they agreed with each one on a scale from strongly disagree (1) to strongly agree (7). Scores for Cognitive Reappraisal (6-items) and Expressive Suppression (4-items) were calculated by following the standard procedure of averaging the items in each category.

Maintenance Tasks Scoring

Affect and Brightness Maintenance Accuracy Scores. The accuracy for each task was determined using the individualized scoring procedure described for Studies 1 and 2. As mentioned above, six participants were excluded from all analysis due to performing two or more standard deviations below the mean on affective or brightness maintenance tasks. These scores were approximately 50% accuracy, indicating chance-level performance.

Results and Discussion

Descriptive statistics for AF, affective maintenance, brightness maintenance, and the VVIQ are provided in Table 2-1. The means make evident that performance on the two maintenance tasks is well matched, and AF performance continued to be relatively stable in this fourth population.

We first examined the correlations among the visual working memory tasks. Brightness maintenance performance did not correlate significantly with change detection performance, $r(82) = 0.13$, $p = .222$, and the relationship with Corsi performance was only a trend, $r(83) =$

0.19, p = .077. In contrast, performance on the Corsi and change detection tasks were significantly related, $r(82) = 0.29$, $p = .008$. These results suggest that the brightness maintenance task measures an ability that is dissimilar from that measured by these two conventional measures of visual working memory. Affect maintenance accuracy did not correlate with either Corsi ($r(83) = .009$, $p = .932$) or change detection performance, $r(82) = -.002$, p = .986. As in Study 2, brightness maintenance accuracy and affect maintenance accuracy were unrelated, $r(83) = 0.12$, $p = 0.258$.

A multiple linear regression was then computed with AF as the outcome variable and performance on affect maintenance, brightness maintenance, Corsi, change detection, and VVIQ as predictors. The model was significant ($F(5, 78) = 2.69$, $p = .027$) with an $R^2 = .15$. Affect maintenance accuracy contributed significantly to the model, predicting AF ability ($\beta = 0.25$, p = .022), while brightness maintenance did not ($\beta = -.03$, $p = .802$; see Figure 2-6). VVIQ scores also contributed significantly, such that greater imaging abilities indicated greater AF accuracy ($\beta = -.21$, $p = .047$). Additionally, performance on the visual change detection task ($\beta = -.22$, p = .049) was also a significant (negative) predictor of AF performance, such that higher capacity scores were predictive of less accurate forecasts. Performance on the Corsi was not a significant predictor of AF, $\beta = .13$, $p = .226$.

The unexpected negative association between change detection performance and AF ability in the multiple linear regression led us to further examine the change detection predictor variable. While a multiple linear regression calculates the variance explained by each predictor holding the other predictors constant, this is only possible when the predictors in the model are orthogonal, or unrelated, to one another (Kutner, Nachtscheim, Neter & Li, 2005). Therefore, correlations greater than zero between predictor variables make it impossible to determine the true unique effect of each predictor on the outcome variable due to multicollinearity, even when the simple correlations between predictors are relatively small (Cohen, Cohen, West & Aiken, 2003). Thus, the significant correlation between Corsi and change detection performance may contribute to this unexpected effect.

To assess the relationship between AF ability and change detection without other predictors, we performed a zero-order correlation and the results were not significant, $r(82) = –.19$, $p = .088$. Because of the difference in predictive value with and without the presence of other variables, it may be that the Corsi, or another variable, acted as a moderator on the relationship between change detection and AF, increasing the effect in its presence (Cohen et al., 2003). Nevertheless, these findings suggest that the change detection variable only predicts AF ability when included in a model with other variables and may not uniquely contribute to AF ability as originally revealed in the regression.

Further support that affect maintenance is uniquely related to AF comes from partial correlations. The relationship between AF and affect maintenance controlling for brightness maintenance was significant ($\rho = .23$, $p =. 036$), whereas the relationship between AF and brightness maintenance controlling for affect maintenance was not ($\rho = –.03$, $p = .789$). The relationship between AF and affect maintenance also remained significant while controlling for Corsi ($\rho = .23$, $p = . 037$) and Change Detection ($\rho = .23$, $p = .035$) performance.

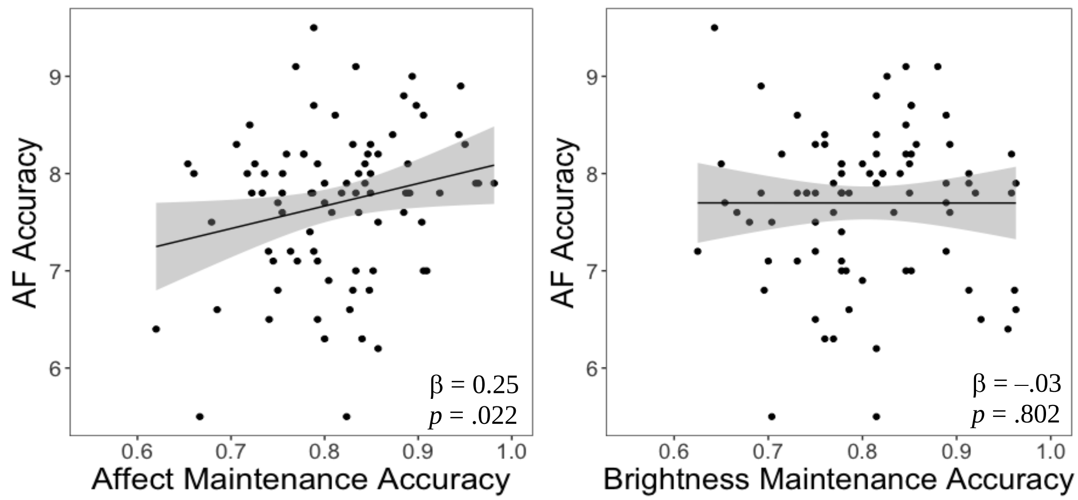

Figure 2-6. Scatterplots showing forecasting accuracy plotted as a function of maintenance task performance in Study 3. The graphs (with best-fitting regression line and shaded region indicating 95% confidence interval) show that Affect Maintenance Accuracy (left) but not Brightness Maintenance Accuracy (right) predicts AF Ability, replicating the findings from Studies 1 and 2.

Exploratory Analyses

To determine whether use of ER strategies were related to AF and affect maintenance performance, Pearson's r correlation coefficients were computed. AF was significantly related to use of cognitive reappraisal ($r(83) = 0.29$, $p = 0.007$) but not expressive suppression ($r(83) = .021$, $p = .845$). Affect maintenance was not significantly related to cognitive reappraisal ($r(83) = .06$, $p = .601$) nor expressive suppression, $r(83) = .09$, $p = .422$.

General Discussion

These studies tested the hypothesis that AWM contributes to the ability to make accurate forecasts about future feelings. Across three independent samples of participants, performance on an affect maintenance task reliably and consistently predicted AF performance, supporting this hypothesis. Furthermore, three studies also demonstrated that working memory for brightness intensity was not predictive of AF, suggesting that working memory for emotion intensity plays a unique role in AF. We consider the implications of these results and several other findings for understanding the processes that underlie AF, as well as the implications for AWM and its role in mental processes more generally. We also consider the finding that EI was not related to AF or AWM. Finally, we discuss several limitations of the present investigations and potential ways to address them.

Implications for Affective Forecasting

We posit that the selective relationship we observed between AWM and AF is due to the specific processes needed to hold in mind and reflect on emotional experiences. This interpretation is supported by the finding that the ability to maintain intensity information about a non-affective subjective state, i.e., brightness, had no bearing on AF ability. Our hypothesis that AWM plays an important role in AF is based on the idea that predicting future feeling states involves conjuring up, working with, and comparing emotional experiences—all processes hypothesized to require AWM (Mikels & Reuter-Lorenz, 2013, 2019; Smith & Lane, 2015). Our results support this hypothesis by demonstrating that individuals who are better able to hold an emotional feeling in mind for subsequent comparison, are better able to predict the intensity of their future feelings.

We included a measure of visual imagery in all studies because a potential strategy for completing the current AF task is to create a visual image of the verbally described emotional scene. It seemed plausible then that people with better and more vivid imagery would perform better on the AF task. Instead, we found that performance on the VVIQ was an inconsistent predictor of AF performance across studies. Combining the data from all four studies revealed no overall relationship between the VVIQ and AF performance ($r(313) = –.06$, $p = .223$), indicating that imagery ability does not facilitate forecasting accuracy at least using these particular measures. Alternatively, the verbal description on each AF trial may be processed directly to evoke the emotional feelings that are then rated for intensity, without the need to generate an image.

For exploratory purposes, the last study included a measure of ER. Whereas the hypothesized relationship with AWM was not evident, people reporting more frequent use of cognitive reappraisal also turned out to be better forecasters. This is an interesting and novel result that requires replication. Nevertheless, the relationship suggests an interaction between forecasting and regulation processes. This is consistent with a proposal by Loewenstein (2007), who posited that participants select ER strategies based on their forecasted effectiveness and that ER success is dependent on the accuracy of these forecasts. Additionally, Ringnes, Stalsett, Hegstad, & Danbolt (2017) discovered that some individuals use prospects of future emotions to regulate their current affective states, an ER strategy they called "emotional forecasting." While the extant findings provide converging evidence for an ER-AF interaction, subsequent research should examine how individuals use their forecasts to inform regulation strategies and vice versa.

Visual working memory did not predict AF performance when measured by the brightness maintenance task across all studies, nor did Corsi Block performance in Study 3. The relationship between AF and change detection is less clear. Better change detection performance was associated with worse forecasting ability in a multiple linear regression but not in an independent correlation. If change detection and AF performance are truly related, it may be that people with lower visual working memory capacities rely on different, more effective, strategies to forecast their feelings. However, given that this relationship was anomalous among the measures of visual working memory used here, and that the negative association between change detection and AF was only significant in presence of other predictors likely reflecting either a moderating effect or collinearity, we are hesitant to conclude that the regression results reflect a true association. Future research using these measures will be needed to resolve these uncertainties.

Implications for Affective Working Memory

The present results indicating that AF is related to affective but not brightness maintenance also provide additional support for the idea that AWM is a separable, domain-specific subsystem of working memory. These results complement and extend prior evidence for separability of affect and brightness maintenance processes based on selective interference methodology, age differences, and special patient populations (Mikels et al., 2008; Mikels et al., 2005, 2012; Gard et al., 2011). Working memory is known to be a fundamental capacity that is integral to higher-order cognition. The present evidence that AWM supports AF ability suggests AWM may play a comparable role as a crucial facet of higher-order emotion-related mentation.

The relationship we found between AWM and AF has practical implications as well. Previous reports suggest that AF may be improved by interventions. For example, by having participants think about multiple past experiences (Buehler & McFarland, 2001) or several features of the target outcome/event when making forecasts (Lam, Buehler, & McFarland, 2005), predictions are less likely to be extreme. The present findings that AWM ability contributes to individual variability in AF suggest that improving AWM could potentially enhance forecasting abilities. Training can improve working memory performance in the cognitive domain (e.g., Soveri, Antfolk, Karlsson, Salo, & Lane, 2017; see also Redick, 2019), and there are indications that the same may be true in the affective domain (e.g., de Voogd, Wiers, Zwitser, & Salemink, 2016). Thus, future work aimed at training AWM could include measures to assess possible benefits to affective forecasting as well.

We also hypothesized AWM and emotion regulation, namely cognitive reappraisal, may share an underlying mechanism due to impaired performance on the affect maintenance task when performing a concurrent cognitive reappraisal task (Mikels et al., 2005; see also Gard et al., 2011). Contrary to our hypothesis, no significant relationship was found between performance on affect maintenance and the self-reported tendency to use cognitive reappraisal. Because the results across the three present studies demonstrate that AWM is predictive of AF ability, and AF ability was found to be related to cognitive reappraisal, it is possible the relationship between AWM and emotion regulation is not direct. Future research should investigate the possible role that AWM may play in cognitive reappraisal and how it may influence ER or other types of higher order emotional processing. Additionally, while a self-report measure for ER is convenient, it may be more informative to examine differences in performance-based cognitive reappraisal success between individuals to determine direct associations.

Relations with Emotional Intelligence

We included EI measures in the current study to reassess the relationships between AF and EI, and to explore the potential relationship between AWM and EI. Across the current studies, we found no evidence for relationships between AF and either trait or ability EI. These results are inconsistent with previous reports that AF is significantly related to ability EI (Dunn et al., 2007; Hoerger et al., 2012) and trait EI (Hoerger et al., 2012). Methodological differences among these studies may contribute to these discrepant results. For both constructs, Dunn et al. (2007) used measures different from those used in the present investigation, measuring AF by comparing predicted and actual reactions to real events (e.g., elections or college sporting events) and measuring ability EI using a composite and emotion management sub-score from the Mayer-

Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey & Caruso, 2002a). These differed from the present verbal-description-based measure of AF and the Situational Test of Emotional Understanding, which focuses on the *understanding emotion* component, rather than the *managing emotion* component of EI. Nevertheless, even using a similar AF measure and identical EI task, we were unable to replicate the Hoerger et al. results (2012). This discrepancy may stem from other prominent differences in study designs. In Hoerger et al. (2012) participants completed the entirety of the AF task online, rather than in-person. Moreover, in Hoerger et al. (Study 2; 2012) participants performed the feelings rating portion of the AF task immediately after the prediction rating with an average delay period of only three minutes, and both used a 9-point rating scale. With this brief intervening interval, memory of prior numerical ratings could influence the assessments and ratings of actual feeling states. In contrast, our participants used a visual analog scale for the rating tasks and had a one-week interval between Phase I (i.e., prediction rating) and Phase II (i.e., experienced feelings rating), design features that are likely to mitigate reliance on memory for the original predictions. These inconsistent findings suggest the need for additional research to understand the relationship between AF and EI.

Moreover, we failed to find associations between EI and AWM. If AWM constitutes a fundamental emotion processing capacity, then one would expect it to be related to both superior EI and more successful AF. However, we found that AWM was unrelated to trait or ability EI. The failure to find relationships between AWM and EI despite finding a consistent relationship between AWM and AF may be due to the differences in the way the constructs were assessed. The measures for AWM and AF were both task-based and used the participants' own responses to determine accuracy. In contrast, both measures of EI are not task-based and accuracy for ability EI using the STEU is determined using Roseman's Appraisal Theory (MacCann &

Roberts, 2008). Future research that tests these associations using another measure such as the MSCEIT, which calculates accuracy using consensus or expert scoring (Mayer et al., 2002a), would allow us to further investigate the generality of these findings.

Limitations and Future Directions

Our study had several limitations that should be addressed in future work. We observed a strong and consistent relationship between AF and AWM. Yet, this association is based on a single measure of AF ability that uses predicted intensity ratings of text describing emotional scenes. While this measure has been used in several prior investigations of AF, an important future direction is to establish that the association we have documented between AWM and AF is generalizable to measures of AF with greater ecological validity. Future studies should investigate this relationship using forecasted feelings for real-life events, which are more likely to approximate the use of AF in everyday life. Our measure of AF was also limited to ratings of anticipated emotional intensity, which aligns well with our intensity-based measure of AWM. A recent model of AF proposed that other aspects of predictions about ones' future feelings, such as the effect of an event on ones' mood, may be more biased, or error prone, than intensity (Lench et al., 2019). We expect that AWM would play a role in these estimates as well, however this remains to be explored.

Other potential limitations are due to the size and age restrictions of our sample. We acknowledge that our sample sizes in each of the four studies were moderate due to practical limitations of in-person laboratory testing. However, our sample sizes were determined by a priori power analyses and are typical compared to other studies using in-person measures of AF (see Buehler & MacFarland, 2001; Dunn et al., 2007; Wilson et al., 1998). In general, in-person testing adds experimental control, consistency, and retention advantages that are less characteristic of internet testing, but also place practical limitations on sample size. Developing an internet version in the future for more wide-spread use of the task would permit a larger and more diverse sample. Additionally, because of age-related effects on emotional processing (Mather, 2016; Mikels et al., 2005), our results may not generalize to other age groups, especially older adults. Given the role of AF in decision making (Loewenstein, 2007; Mellers & McGraw, 2001) and the life-altering financial and healthcare decisions faced by older adults, future studies should take a life-span developmental approach to determine if and when the relationship between AWM and AF changes.

Finally, a potential caveat concerns the concurrent validity of the brightness maintenance task. The demands of the brightness maintenance task were well-suited to juxtapose the affect maintenance task, where the emotional *intensity* attribute is held in working memory. However, brightness maintenance was only weakly and non-significantly related to Corsi and change detection performance. The limited associations between brightness maintenance and these two canonical visual working memory tasks could reflect that the brightness maintenance task measures the ability to accurately retain an attribute of a visual image, namely *intensity*, rather than quantities or sequences of visual features (e.g., colors or locations). Thus, while the concurrent validity of the brightness maintenance task remains to be further clarified, affect and brightness maintenance abilities do appear separable (as evidenced by their lack of correlation and the selective relationship between AF and maintenance of emotional, but not brightness, intensity). Furthermore, the relationship between affect maintenance and AF remained unaffected when controlling for brightness maintenance and either of the two canonical measures of visual working memory.

Conclusion

In conclusion, despite considerable extant research on AF, the contributions of an elemental system such as AWM have not been previously considered. The present results indicate that unlike other measures of working memory, affect maintenance ability reliably predicts AF ability. We conclude that AWM is a core ability that supports affective prospection, and may also underlie other forms of higher-order emotional thought. For example, AWM may provide the mental work space for episodic counterfactual thinking about past emotional experiences (De Brigard & Parikh, 2019), or other forms of mental simulation where feeling states play a prominent role. The present work suggests that individual differences in AWM ability may be important to the quality and accuracy of emotion-related thought.

Context

Affective working memory (AWM) is a putative core mental ability for higher-order emotion-related thought. Because affective forecasting (AF) involves judgements about feelings that are generated, actively maintained in mind and evaluated, processes that we and others attribute AWM, we hypothesized that AWM ability would contribute to the accuracy of affective forecasts. That is, if someone is better able to conjure and hold emotional feeling states in mind, they would be more accurate in reflecting on and predicting their future feelings. While much of the affective forecasting literature focuses on errors due to bias, our work focuses on individual differences due to working memory abilities. The current findings support the hypothesis that AWM plays a unique role in AF ability adding credence to the proposal that AWM contributes to higher-order emotional thought. In future work, we plan to assess this relationship further using other forecasting tasks, and extend this research to older adults to determine how the AWM-AF relationship changes with age.

References

Andrews, F. M., & Robinson, J. P. (1991). Measures of subjective well-being. *Measures of Personality and Social Psychological Attitudes*, *1*, 61–114. http://dx.doi.org/10.1016/B978-0-12-590241-0.50007-1

Baddeley, A. (1992). Working memory. *Science*, *255*(5044). 556–559. http://dx.doi.org/10.1126/science.1736359

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. http://dx.doi.org/10.1146/annurev-psych-120710-100422

Broome, R., Gard, D. E., & Mikels, J. A. (2012). Test–retest reliability of an emotion maintenance task. *Cognition & Emotion*, *26*(4), 737–747. http://dx.doi.org/10.1080/02699931.2011.613916

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin*, *27*(11), 1480–1493. http://dx.doi.org/10.1177/01461672012711009

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates, Inc.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Dissertation Abstracts International, 34, 819B.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–114. https://doi.org/10.1017/S0140525X01003922

De Brigard, F., & Parikh, N. (2019). Episodic counterfactual thinking. *Current Directions in Psychological Science*, *28*(1), 59-66. http://dx.doi.org/10.1177/0963721418806512

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313*(5787), 684–687. http://dx.doi.org/10.1126/science.1128356

de Voogd, E. L., Wiers, R. W., Zwitser, R. J., & Salemink, E. (2016). Emotional working memory training as an online intervention for adolescent anxiety and depression: A randomised controlled trial. *Australian Journal of Psychology, 68,* 228 –238. http://dx.doi.org/10.1111/ajpy.12134

Dunn, E.W., Brackett, M. A., Aston-James, C., Schneiderman, E., & Salovey, P. (2007). On emotionally intelligent time travel: Individual differences in affective forecasting ability.

Personality and Social Psychology Bulletin, *33*(1), 85–93. http://dx.doi.org/10.1177/0146167206294201

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Gard, D. E., Cooper, S., Fisher, M., Genevsky, A., Mikels, J. A., & Vinogradov, S. (2011). Evidence for an emotion maintenance deficit in schizophrenia. *Psychiatry Research*, *187*(1), 24–29. http://dx.doi.org/10.1016/j.psychres.2010.12.018

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*(3), 617-638. http://dx.doi.org/10.1037/0022-3514.75.3.617

Gross, J. J. (2013). Emotion regulation: Conceptual and empirical foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed., pp. 1–20). New York: The Guilford Press.

Gross, J.J., & John, O.P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85,* 348-362. http://dx.doi.org/10.1037/0022-3514.85.2.348

Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, *8*(1), 35-41. http://dx.doi.org/10.1111/j.1745-3992.1989.tb00313.x

Hoerger, M., Chapman, B. P., & Duberstein, P. R. (2016). Realistic affective forecasting: The role of personality. *Cognition and Emotion*, *30*(7), 1304–1316. http://dx.doi.org/10.1080/02699931.2015.1061481

Hoerger, M., Chapman, B. P., Epstein, R., & Duberstein, P. R. (2012). Emotional intelligence: A theoretical framework for individual differences in affective forecasting. *Emotion*, *12*(4), 716–725. http://dx.doi.org/10.1037/a0026724

Hughes, D. J. & Evans, T. R. (2018). Putting ‘emotional intelligences’ in their place: Introducing the integrated model of affect-related individual differences. *Frontiers in Psychology, 9*, 1–15. http://dx.doi.org/10.3389/fpsyg.2018.02155

JASP team. JASP (Version 0.9) [Computer software]. Retrieved from https://jasp-stats.org/previous-versions/

Kutner, M. H., Nachtscheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th Edition). New York: McGraw-Hill.

Lam, K. C., Buehler, R., McFarland, C., Ross, M., & Cheung, I. (2005). Cultural differences in affective forecasting: The role of focalism. *Personality and Social Psychology Bulletin, 31*(9), 1296–1309. http://dx.doi.org/10.1177/0146167205274691

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. Gainesville: University of Florida, Center for Research in Psychophysiology.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016-E2025. http://dx.doi.org/10.1073/pnas.1619316114

Lench, H. C., Levine, L. J., Perez, K., Carpenter, Z. K., Carlson, S. J., Bench, S. W., & Wan, Y. (2019). When and why people misestimate future feelings: Identifying strengths and weaknesses in affective forecasting. *Journal of Personality and Social Psychology, 116*(5), 724-742. http://dx.doi.org/10.1037/pspa0000143

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology, 103*(4), 584–605. http://dx.doi.org/10.1037/a0029544

Loewenstein, G. (2007). Affect regulation and affective forecasting. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 180-203). New York, NY: Guilford Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature, 390, 279–281. https://doi.org/10.1038/36846

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*(4), 540–551. http://dx.doi.org/10.1037/a0012746

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. http://dx.doi.org/10.1111/j.2044-8295.1973.tb01322.x

Mather, M. (2016). The affective neuroscience of aging. *Annual Review of Psychology, 67,* 213–238. http://dx.doi.org/10.1146/annurev-psych-122414-033540

Mathieu, M. T., & Gosling, S. D. (2012). The accuracy or inaccuracy of affective forecasts depends on how accuracy is indexed: A meta-analysis of past studies. *Psychological Science, 23*(2), 161-162. http://dx.doi.org/10.1177/0956797611427044

Mayer, J. D., Salovey, P., & Caruso, D. (2002a). *The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), Version 2.0*. Toronto, Canada: Multi-Health Systems.

Mayer, J., Salovey, P., & Caruso, D. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*, 197–215. http://dx.doi.org/10.1207/s15327965pli1503_02

Mellers, B. A., & Mcgraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10*(6), 210–214. http://dx.doi.org/10.1111/1467-8721.00151

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging*, *20*(4), 542–553. http://dx.doi.org/10.1037/0882-7974.20.4.542

Mikels, J. A., Reed, A. E. (2009). Monetary loses do not loom larger later in life: Age differences in the framing effect. *Journal of Gerontology: Psychological Sciences, 64B*(4), 457–460. http://dx.doi.org/10.1093/geronb/gbp043

Mikels, J. A., & Reuter-Lorenz, P. A. (2013). Emotion and working memory. In H. Pashler (Ed.), *Encyclopedia of the mind* (pp. 308–310). Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781452257044.n113

Mikels, J. A., Reuter-Lorenz, P. A. (2019). Affective working memory: An integrative psychological construct. *Perspectives in Psychological Science*, 1–17. http://dx.doi.org/10.1177/1745691619837597

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion*, *8*(2), 256–266. http://dx.doi.org/10.1037/1528-3542.8.2.256

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15*, 425–448. https://doi.org/10.1002/per.416

Redick, T. S. (2019). The hype cycle of working memory training. *Current Directions in Psychological Science, 28*(5), 423–429. https://doi.org/10.1177/0963721419848668

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*(1), 5–21. http://dx.doi.org/10.1016/j.neuroscience.2005.12.061

Ringnes, H. K., Stalsett, G., Hegstad, H., & Danbolt, L. J. (2017). Emotional forecasting of happiness. *Archive for the Psychology of Religion, 39*, 312–343. http://dx.doi.org/10.1163/15736121-12341341

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin*, *27*(11), 1520–1532. http://dx.doi.org/10.1177/01461672012711012

Scheibe, S., Mata, R., & Carstensen, L. L. (2011). Age differences in affective forecasting and experienced emotion surrounding the 2008 US presidential election. *Cognition & Emotion*, *25*(6), 1029–1044. http://dx.doi.org/10.1080/02699931.2010.545543

Schutte, N. S., Malouff, J. M., & Bhullar, N. (2009). The assessing emotions scale. In *Assessing emotional intelligence* (pp. 119-134). Boston, MA: Springer. http://dx.doi.org/10.1007/978-0-387-88370-0_7

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, *25*(2), 167–177. http://dx.doi.org/10.1016/S0191-8869(98)00001-4

Smith, E. E. & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657–1661. http://dx.doi.org/10.1126/science.283.5408.1657

Smith, R., & Lane, R. D. (2015). The neural basis of one's own conscious and unconscious emotional states. *Neuroscience & Biobehavioral Reviews*, *57*, 1-29. http://dx.doi.org/10.1016/j.neubiorev.2015.08.003

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*(4), 1077–1096. http://dx.doi.org/10.3758/s13423-016-1217-0

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, *35*, 345–411. http://dx.doi.org/10.1016/S0065-2601(03)01006-2

Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology*, *105*, 740–748. https://doi.org/10.1037/a0032662

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78*(5), 821–836. http://dx.doi.org/10.1037/0022-3514.78.5.821

Chapter 3 The Role of Affective Working Memory in Real-World Affective Forecasting Accuracy (Study 4)

Choice is a fundamental part of life—Should I ride my bike to work today? Should I take that new job across the country? Should I splurge on a fancy television? And a crucial aspect of making good decisions is predicting how different outcomes will make us feel (Mellers, Schwartz, & Ritov 1999; Loewenstein & Lerner, 2003; Charpentier et al., 2016). However, this form of prospection, known as affective forecasting (Wilson and Gilbert, 2003), is an ability that varies greatly amongst individuals. A recent investigation sought to explain variability in affective forecasting and found that *affective working memory*—a distinct domain of working memory responsible for actively maintaining and working with feeling states—predicted individual differences in forecasting accuracy (Frank et al., 2021). In the present study, we tested whether this relationship between the ability to maintain emotional experiences and forecasting accuracy, previously tested in the laboratory, generalized to forecasted feelings to a major real-world event (i.e., 2020 United States presidential election) in a broad sample of participants recruited via the internet. This work aimed to expound the unique role affective working memory plays in emotional prospection.

Affective working memory, also called working memory for emotion, refers to the ability to actively hold feeling states in mind (Davidson & Irwin, 1999; Mikels et al., 2005; 2008; Mikels & Reuter-Lorenz, 2019). Just as cognitive working memory supports the maintenance of mental representations of perceptual or verbal information, affective working memory constitutes a distinct working memory domain responsible for the maintenance of emotion representations where feelings themselves are the memoranda (Mikels et al., 2008; Mikels & Reuter-Lorenz, 2019). Affective working memory is typically measured using an emotion maintenance (also called affect maintenance) task, where participants hold feelings evoked by an emotional image in mind over a delay before deciding if the feelings were more or less intense than feelings elicited by a second image (Mikels et al., 2005; 2008; Broome et al., 2012). Although separable from other working memory domains, affective working memory also reflects an active, rather than passive, maintenance process (Waugh, Lemus, & Gotlib, 2014; Smith et al., 2017, DeFraine, 2016), and has been posited to support goal-directed behavior across several higher-order emotion-related abilities (e.g., emotion regulation, wisdom, rumination; Mikels & Reuter-Lorenz, 2019). Most recently, affective working memory was found to explain variability in the accuracy of forecasted feelings (Frank et al., 2021).

According to decision affect theory, we use predictions of future feelings (i.e., affective forecasting) to guide our choices (Mellers, Schwartz, & Ritov, 1999; Mellers & McGraw, 2001). Several studies have found that incorporating anticipated emotions into decision-making models better predicted choice, underscoring the idea that affective forecasting is an important part of how we make decisions (Mellers et al., 1999; Ahn et al., 2012; Charpentier et al., 2016; Hayes & Wedell, 2020). Yet, we are not all that accurate in these predictions, often overestimating the intensity and duration of our feelings (i.e., impact bias; Wilson & Gilbert, 2003; 2013; but see Levine et al., 2019), which may lead to suboptimal decision making (Kermer et al., 2006; Loewenstein, 2005). Inaccurate prospection, including affective predictions, has also been associated with negative mental health and well-being outcomes (Gilbert & Wilson, 2007; MacLeod, 2016; Roepke & Seligman, 2016). Although this ability to make accurate predictions of future feelings varies greatly among individuals, the reasons for this variability are poorly understood. Some previous studies have proposed that individual differences in forecasting accuracy may be related to personality traits (Hoerger et al., 2016) or emotional intelligence (Dunn et al, 2007; Hoerger et al., 2012; but see Frank et al., 2021). Alternatively, in a recent study, Frank et al. (2021) found that variability in forecasting accuracy may also be explained by working memory for emotion—i.e., *affective working memory*. Because affective forecasting requires conjuring up and maintaining emotional experiences for evaluation, the authors hypothesized that individuals who are better able to maintain feeling states (i.e., better affective working memory), would be more accurate in predicting their future feelings. Across a series of three laboratory-based studies, emotion, but not perceptual, maintenance abilities predicted forecasting accuracy, revealing a unique role for affective working memory in the accuracy of anticipating future feelings (Frank et al., 2021).

The measure of affective forecasting accuracy used by Frank et al. (2021), employed previously by Hoerger et al. (2012), was a description-based task that compared predicted and actual feelings elicited by emotionally evocative photographs. If affective working memory plays a unique and fundamental role in emotional prospection, then individual differences in this type of working memory should also predict the accuracy of forecasts about real-life events. The present study tested this hypothesis using a major real-world event: the 2020 United States presidential election. Further, this study included a replication of the description-based forecasting task where the data were obtained from a broad internet-based participant sample. As hypothesized, we report that emotion, but not perceptual, maintenance abilities predict individual differences in forecasting accuracy for both a major real-life event and the description-based task.

Method

Participants

A total of 114 participants were recruited—110 from Prolific, an online recruitment platform, and four[1] from the University Introductory Psychology Subject Pool. Ninety-four of these participants returned to complete Session 2, and 86 completed Session 3. Following a priori exclusion criteria, ten outliers were removed for poor performance (see Design and Procedure for details). Thus, the final sample included 76 participants—73 from Prolific and three from Subject Pool[2]. These 76 participants (aged 18-29, mean age = 23.6, 51.3% female, 56.6% students) were self-reported right-handed, English-speakers. Because our key measure incorporates the outcome of the United States (U.S.) presidential election, only U.S. citizens were recruited. We aimed to obtain a sample size of at least $N = 85$ to match previous studies that used a priori power analyses to examine similar measures of interest (Frank et al., 2021). Although this was ultimately not possible due to high rates of attrition, our actual sample of 76 participants led to the observed power ($1 - \beta$) of 0.72 and 0.73 for each model, which was sufficient for detecting small to medium effects (Cohen's f = .123 and .135; Prajapati, Dunne & Armstrong, 2010)[3]. All experimental procedures were approved by the University of Michigan Institutional Review Board. The methods and analyses were preregistered and can be found at https://aspredicted.org/blind.php?x=et412p.

---

[1] Due to a rise in Covid-19 cases on campus, a stay-at-home order for undergraduate students was issued just two days before data collection began, which likely contributed to the few number of Subject Pool participants.
[2] Too few participants were recruited from Subject Pool to make between-samples comparisons, however, our findings hold whether or not these three participants are included.
[3] Issues regarding the sample size are further addressed in the discussion.

Design and Procedure

Participants completed three testing sessions spanning a three-week period (see Figure 3-1). In Session 1, participants completed the brightness maintenance and brightness intensity rating tasks, and Phase I of the description-based forecasting measure. In Session 2, one-week later, participants completed Phase II of the description-based measure, the emotion maintenance and emotion intensity rating tasks, and Phase I of the election outcome forecasting measure. During Session 3, participants completed Phase II of the election outcome forecasting task. Session 3 took place two weeks after Session 2, approximately ten days after the election, and six days after the election was called by official sources. The maintenance task order and description-based forecasting delay was intended to match Frank et al. (2021). The timing and lag used for the election outcome was informed by previous studies that used periods ranging from a few days to more than a month before and after the election (Meyvis, Ratner, & Levav, 2010; Emmanuel et al., 2010; Lench et al., 2019). Maintenance and rating tasks were created in PsychoPy (Version 2020.1.2; Peirce et al., 2019) and hosted via Pavlovia. Forecasting measures were created and completed on Qualtrics. Participants completed the three study sessions online using their own personal computers. During sessions 1 and 3, participants also completed a novel measure of forecasting using real-life events, which will not be discussed further (a fuller description of this measure and the findings are available in the Appendix 3A).

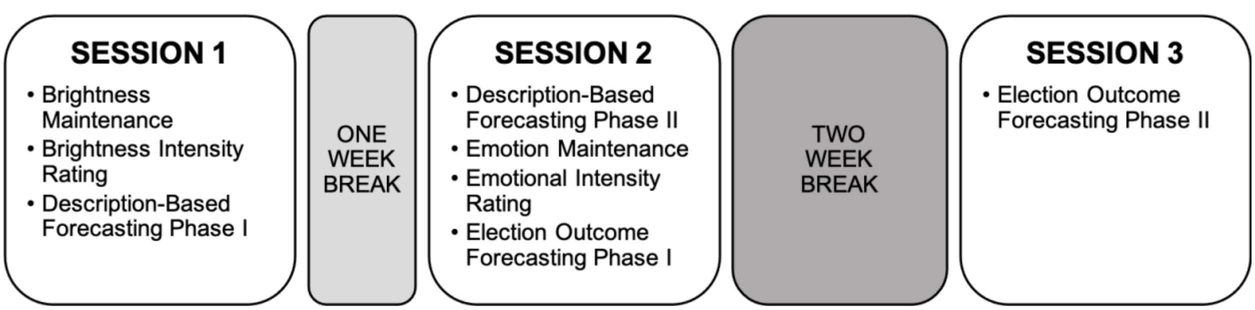

Figure 3-1. Protocol for Study 4.

Emotion Maintenance. Affective working memory was assessed with an emotion maintenance task (Figure 3-2; Mikels et al., 2005; 2008; Broome et al., 2012; Frank et al., 2021). For each trial, participants viewed one emotional image (5 seconds). After it disappeared, participants continued to actively hold the feelings elicited from the photo, at the same intensity level (3 seconds). They then viewed another emotional image (5 seconds), before deciding if the second image evoked more or less 'emotional intensity' (i.e., strength or amount of emotional reaction), compared to the first. Participants made their responses using the right (second image evoked more intensity) or left (second image evoked less intensity) arrow keys. Each of the 56 trials consisted of a matched-valence image pair (28 positive, 28 negative), with images taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1999). Pairs were the same as in Frank et al. Study 3 (2021), which used an item-response theory (IRT) analysis designed to identify pairs that were most discriminative between high and low performers, and matched mean accuracy scores with the brightness maintenance task (see Frank et al. 2021 Supplemental Method for full details).

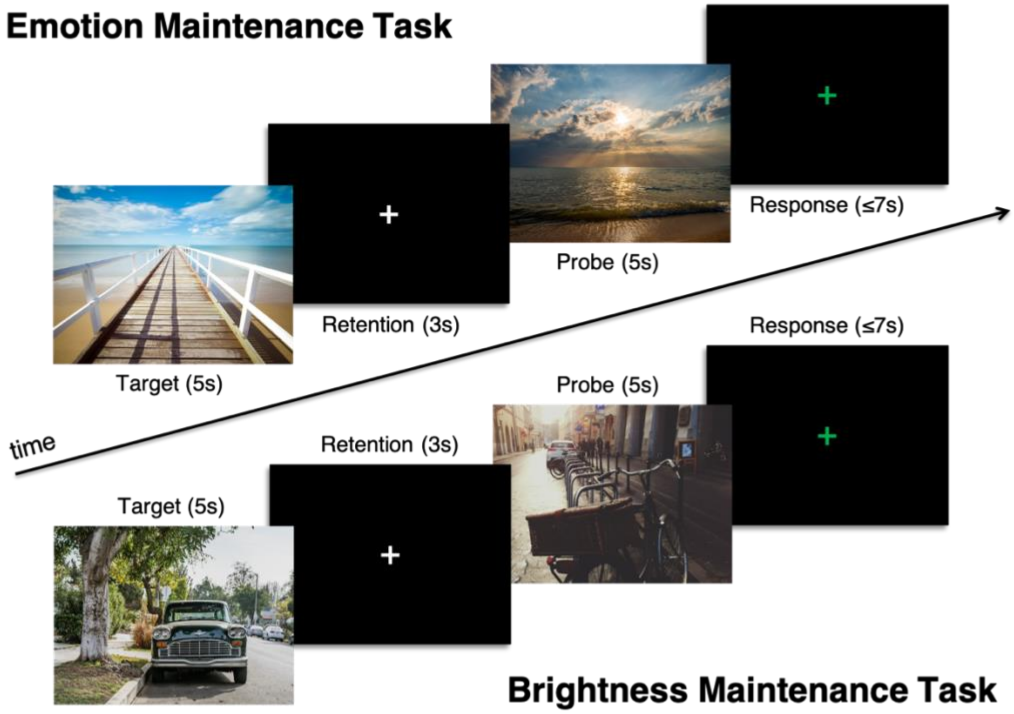

Figure 3-2. Schematic of maintenance tasks used in Study 4. Participants hold the emotion or brightness intensity of one image in mind over a delay, before determining if a second image has more or less intensity. Image adapted from "Affective Working Memory: An Integrative Psychological Construct" by J. A. Mikels and P. A. Reuter-Lorenz, 2019, Perspectives in Psychological Science, 14(4), p. 8. Copyright 2019 by SAGE Publications, also used in Frank et al. (2021).

Emotion Intensity Rating. Participants viewed the 112 images from the emotion maintenance task again individually. For each image, participants rated the emotion intensity (i.e., strength or amount of emotional reaction) elicited from the image using a visual analog scale with anchors *not at all intense* and *extremely intense*. Participants clicked anywhere along the continuous scale (absolute resolution: 21 points) to indicate the degree of emotion intensity for the image. These ratings were then used to calculate individualized accuracy scores for the emotion maintenance task (see Maintenance Task Scoring).

Brightness Maintenance. Non-affective maintenance abilities were assessed with a brightness maintenance task (Figure 3-1), designed as an emotionally-neutral analogue task requiring intensity maintenance (Mikels et al., 2005; 2008; Broome et al., 2012). For each trial, participants viewed one neutral image (5 seconds). After it disappeared, participants continued to actively hold the brightness intensity in mind, at the same level (3 seconds). They then viewed a second neutral image (5 seconds), before deciding if the second image evoked more or less 'brightness intensity' (i.e., amount of overall light or illumination), compared to the first. Participants used the left and right arrow key to make their selection. There were 28 pairs in all (56 images) selected by Frank et al. Study 3 (2021).

Brightness Intensity Rating. Participants viewed the 56 images from the brightness maintenance task again individually and rated the brightness intensity of each one on a visual analog scale (absolute resolution: 21 points) ranging from *not at all intense* to *extremely intense*. These ratings were used to calculate individualized accuracy scores for the brightness maintenance task (see Maintenance Task Scoring).

Maintenance Task Scoring. To calculate individualized accuracy scores for the emotion maintenance and brightness maintenance tasks, we used the intensity ratings to infer which image, when appearing with its pair during the maintenance task, should have been judged as evoking more intensity. A trial was scored as correct if there was agreement between the intensity ratings and maintenance responses (e.g., second image rated higher during intensity rating task and more intense during maintenance task), or incorrect if they were inconsistent (e.g., second image rated higher during intensity rating task but less intense during maintenance task). Trials in which participants rated both images in a pair as having the same intensity were excluded, accounting for 8–9% of trials. We then averaged the scores across all trials, separately for each maintenance task, to obtain emotion-maintenance and brightness-maintenance accuracy scores between 0 and 1. As described in the pre-registration, we excluded participants who performed > 2 standard deviations below the mean. This resulted in removal of six participants for the brightness maintenance task and zero for the emotion maintenance task.

Description-Based Forecasting. To measure affective forecasting ability as tested previously by Frank et al. (2021), we used a description-based measure adapted from Hoerger et al. (2012). In Phase I (Session 1), participants read descriptions of ten emotional scenes (five positive, five negative) and predicted how they would feel if they were to view each image using a visual analog scale ranging from *extremely unpleasant* to *extremely pleasant*. In Phase II (Session 2), participants viewed each emotionally evocative photograph and rated how they felt using the same scale (absolute resolution: 21 points). We then calculated the absolute difference between ratings in Phase I and Phase II, averaged across all images, to obtain a forecasting error score for each participant. The maximum possible error was 20, with scores > 10 indicating a valence reversal (e.g., negative prediction but positive experience). Following the pre-registered exclusion criteria, one outlier (forecasting error > 3 SD above the mean) was removed.

Election Outcome Forecasting. To measure affective forecasting that better captures real-life affective predictions, we adapted an event-based forecasting measure (Lench et al., 2019) to compare predicted and experienced feelings to the outcome of the 2020 U.S. presidential election. In Phase I (Session 2), participants read "Suppose it is an evening in November and (Joe Biden/Donald Trump) has won the election and will be the next president of the United States. How will you be feeling about (Joe Biden/Donald Trump) being president?" Participants then rated their predictions of feeling happy, angry, and scared for both candidates on a visual analog scale (absolute resolution: 21 points) with anchors *not at all (happy/angry/scared)* and *extremely (happy/angry/scared),* taken from Lench et al. (2019). In Phase II (Session 3), approximately ten days after the election, and six days after the election was called by official sources, participants reported their experienced feelings for the outcome that occurred (i.e., Biden victory) using the same three scales. We then calculated the absolute discrepancy between ratings in Phase I and Phase II, averaged across the three emotions, resulting in a composite forecasting error score (or “forecasting error”) for each participant with possible values ranging between 0 and 20. Following the a priori exclusion criteria, one outlier was removed (forecasting error > 3 SD above the mean). Additionally, due to computer error, data from one participant was not recorded on this measure.

Results

Descriptive Statistics

Mean accuracy scores for the emotion maintenance and brightness maintenance tasks, and mean forecasting error for the description-based and election outcome measures, are provided in Table 3-1. As can be seen from this table, brightness-maintenance accuracy (M = .75, SD = .11) was significantly higher than emotion-maintenance accuracy (M = .69, SD = .15, $t(75) = 3.25$, $p = .002$). Forecasting error on the description-based measure was consistent with findings from Frank et al. (2021), and did not differ from election outcome forecasting error, $t(82) = .92$, p =.36.

Table 3-1. Mean (SD) Performance Across Measures in Study 4

Measure	Mean (SD)
Emotion-Maintenance Accuracy	0.69 (.15)
Brightness-Maintenance Accuracy	0.75 (.11)
Description-Based Forecasting Error	2.59 (1.04)
Election Outcome Forecasting Error	2.81 (2.22)

Note: Maintenance task accuracy scores range from 0 and 1; Forecasting error scores range from 0 – 20, with higher scores reflecting less accurate predictions

Primary Analyses

To test whether maintenance performance predicts forecasting error across the three measures, we ran a multivariate linear regression with description-based and election outcome forecasting error as the outcome variables and emotion-maintenance and brightness-maintenance performance as the predictors. We also included two covariates to control for individual variations in the delay between Sessions 1 and 3, and between Sessions 2 and 3, respectively[4]. Following our pre-registered analyses, we also included an exploratory forecasting measure using personal life events as a third outcome variable, which yielded null results (this measure and findings are described in full in Appendix 3A).

Emotion-maintenance performance significantly predicted forecasting error on both the description-based (β = –.292, p = .023; Figure 3-3a) and election outcome (β = –.33, p = .009; Figure 3-3b) measures such that better ability to maintain emotions predicted more accurate forecasting both in the laboratory and real-life. By comparison, brightness-maintenance performance did not predict forecasting variability for either the description-based (β = –.084, p = .503) or election outcome (β = .177, p = .149) measure.

Relationship Between Forecasting Measures

We also examined the association between the description-based and election outcome forecasting measures using a Pearson's correlation coefficient. The association between forecasting errors on the description-based and election outcome measures did not reach significance, $r(79) = .13$, $p = .248$.

---

[4] Modifying the covariates to account for delay between Sessions 1 and 2 did not alter the findings.

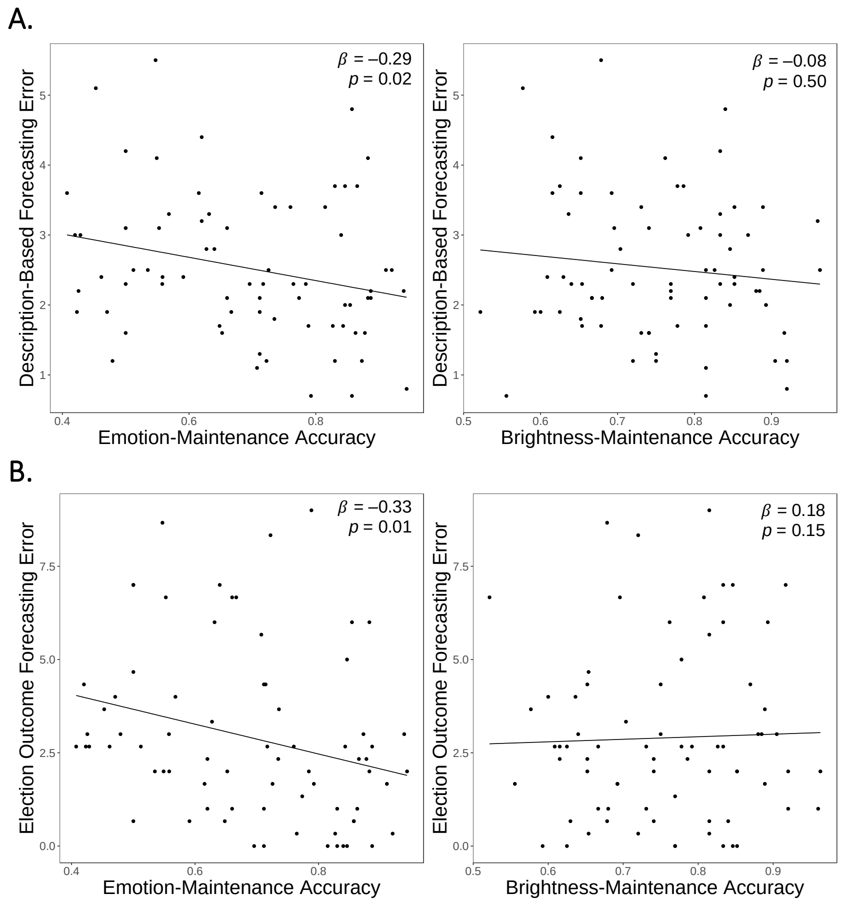

Figure 3-3. Scatterplots showing forecasting error plotted as a function of maintenance task performance in Study 4. The graphs (with best-fitting regression line, shaded regions reflecting 95% confidence intervals, standardized betas and p-values in left-hand corner) reveal that emotion maintenance (left) predicts forecasting error on the description-based (A) and election outcome (B) measures, whereas brightness maintenance (right) does not.

Valence-Specific Analyses

To determine whether the relationship between maintenance task performance and forecasting accuracy was valence-specific, we conducted additional analyses using only positive or only negative trials. Similar to what was found in Frank et al. (2021), no significant associations emerged from these analyses, suggesting that the relationship between overall forecasting accuracy and emotion maintenance is not valence-specific.

Direction of Forecasting Bias

To assess whether participants showed a tendency to overestimate or underestimate their future feelings, we also calculated a forecasting bias score (i.e., the raw difference between experienced and predicted feelings) for each participant. Based on the scales used here, a positive bias score (i.e., greater than zero) would reflect a participant feeling better than expected, that is, an overestimation of negative feelings or an underestimation of positive feelings. Accordingly, a negative bias score (i.e., less than zero) would reflect a participant feeling worse than expected, signifying an underestimation of negative feelings or an overestimation of positive feelings. Additionally, because misestimations may be valence-dependent (see Wilson & Gilbert, 2003 for a review), we followed the standard practice used in the forecasting literature of assessing bias separately for positive and negative outcomes (see Table 3-2).

For the description-based forecasting measure, we calculated forecasting bias separately for positive and negative trials—i.e., the valence of each emotionally evocative image. For positive trials, participants significantly overestimated their positive feelings, predicting they would feel better than actually experienced (p = .001; see Table 3-2 for results). For negative trials, participants overestimated the degree of their negative feelings, expecting to feel worse than experienced (p < .0001).

To assess forecasting bias on the election outcome measure, we categorized participants into valence-specific groups based on whether they considered the Biden victory to be a positive or negative outcome. Sixty-seven participants were classified based on their reported preferred candidate (Biden: Positive outcome, Trump: Negative outcome). We then classified the remaining 17 participants who did not select either major-party nominee as their preferred candidate into positive or negative outcome groups based on the valence of their predicted feelings of a Biden victory. Participants who saw the Biden victory as a positive outcome did not show an overall forecasting bias (*p* = .683; see Table 3-2). However, those who saw the Biden victory as a negative outcome overestimated how negative they would feel (*p* = .002). This pattern remains true if we only include the 67 participants who selected either Biden or Trump as their preferred candidate (Biden-supporters, *p* = .634; Trump-supporters, *p* = .044).

Table 3-2. Mean (SD) Forecasting Bias by Valence in Study 4

Forecasting Measure	Valence	N	Raw Forecasting Error Mean (SD)	Bias Direction
Description-Based	Positive	83	–0.93 (2.46)	*p* = .001; Overestimation
	Negative	83	0.86 (1.52)	*p* < .0001; Overestimation
Election Outcome	Positive	66	0.14 (2.80)	*p* = .683; None
	Negative	18	2.67 (3.04)	*p* = .002; Overestimation

Discussion

The present study provides new evidence that affective working memory is a core ability that affects the accuracy with which people can forecast their future feelings. Performance on a task requiring active maintenance of emotion intensity predicted individual differences in description-based forecasting accuracy in a broader sample of participants aged 18-29, living across the United States. In contrast, the ability to maintain brightness intensity did not. This selective relationship was further replicated when forecasting feelings about a major real-world event (i.e., 2020 U.S. presidential election). We consider the implications of these results for affective working memory and affective forecasting, including directions for future research and the limitations of this work.

As hypothesized, better emotion-maintenance performance predicted more accurate forecasting using a description-based measure. This result replicates and extends our previous report of a selective positive relationship between affective, but not perceptual or cognitive, working memory and forecasting accuracy (Frank et al., 2021). In Frank et al. (2021), participants—college students attending a large, public, midwestern university—were tested in person in the laboratory. In the present study, young adult participants completed the same measures using an online format. Participants in this sample were made up of only 56% students and came from 32 states within the U.S. Demonstrating this pattern with on-line testing in a broader sample attests to its robustness and generalizability.

Emotion-maintenance performance also predicted the accuracy of feelings forecasted in response to the outcome of the U.S. presidential election, a relationship not previously assessed. This result confirms our hypothesis that affective working memory plays an important role in forecasting abilities including emotional prospection about highly impactful real-world outcomes. In contrast, brightness-maintenance performance did not predict forecasting error in either the description-based or election outcome forecasting measures. This result corroborates previous findings that brightness maintenance and two other types of perceptual working memory did not independently predict variability in the description-based forecasting accuracy (Frank et al., 2021). Therefore, it appears that cognitive working memory does not account for individual differences in forecasting accuracy, providing further support for a unique role of affective working memory in emotional prospection.

Implications & Future Directions

These findings advance our understanding of affective working memory in several important ways. Affective working memory has been posited to be related to a variety of higher-order emotion-related mental abilities including emotion regulation, rumination (Mikels & Reuter-Lorenz, 2013; 2019) and most recently, affective forecasting (Frank et al., 2021). The present evidence reveals a direct association between affective working memory and real-world forecasting—an form of prospective thought that is vital to decision making and quality of life. These results demonstrate the importance of affective working memory as a fundamental psychological process that contributes to higher-order emotional thought. Just as cognitive working memory provides the mental workspace for complex cognitive tasks, affective working memory provides the mental workspace for mental simulations where feelings play a prominent role. Accordingly, affective working memory may support other types of emotional prospection, such as counterfactual thinking—or other forms of mental simulation that involve working with emotions.

These results also have important implications for understanding the psychological abilities that contribute to affective forecasting. Existing models of forecasting focus on the sources of error and bias (e.g., Wilson & Gilbert, 2003; Lench et al., 2019) and point to the need for interventions aimed at cognitive strategies to overcome these forecasting errors (Buehler & McFarland, 2001; Roepke & Seligman, 2016). Our results underscore the important role of fundamental psychological abilities, like affective working memory, in forecasting accuracy. This idea may lead to updated forecasting models that focus on how our core mental processes contribute to the accurate predictions of future feelings. Moreover, to better understand the role of working memory for emotion in emotional prospection, future research should work to identify which types of forecasting errors are most closely tied to affective working memory. It may be the case that affective working memory is strongly related to the *projection bias*—i.e., the tendency for our current affective states to influence our predicted feelings (Loewenstein et al. 2001; Loewenstein, O'Donoghue, & Rabin, 2003; Wilson & Gilbert, 2005)—such that those with a worse ability to hold both current and predicted emotions in mind are more likely to demonstrate this bias, compared to those with higher affective working memory abilities. On the other hand, emotion-maintenance abilities might play a smaller role in errors such as *misconstrual*—i.e., imagining an inaccurate future event (i.e., Wilson & Gilbert, 2003). Finding the errors in which affective working memory plays the largest role will provide a better understanding how individuals forecast their feelings, and allow researchers to create the most effective tools for enhancing these abilities.

Limitations

One potential limitation of the present study is the overall level of maintenance task performance, which was lower than obtained with in-person[5] testing (Frank et al. Study 3, 2021). We attribute this lower accuracy to online data collection. Participants recruited through online platforms have been shown to display lower motivation and/or effort (Chmielewski & Kucker, 2020; Newman et al., 2021), which may be especially consequential for mentally demanding tasks. Nonetheless, even after using more stringent performance cutoffs to eliminate participants who may be performing at chance-level (N = 37)[6], the novel relationship between emotion maintenance and election outcome forecasting accuracy remains significant. This is also the case if we exclude participants based on the maintenance task cutoffs obtained from in-person testing (Frank et al. Study 3, 2021). Thus, it is unlikely that our present results can be explained by poor maintenance task performance.

---

[5] Internet-Based Testing Accuracy (current study): Emotion Maintenance: M = 0.69, Brightness Maintenance: M = 0.75; In-Person Testing Accuracy (Frank et al. Study 3, 2021): Emotion Maintenance: M = 0.80, Brightness Maintenance: M = 0.81.

[6] To identify a range of chance performance, we created 95% confidence intervals using parametric bootstrapping (10,000 iterations), completed separately for emotion (56 trials) and brightness (28 trials) maintenance tasks.

Furthermore, participants performed better on the brightness, compared to the emotion, maintenance tasks. We believe the issues with online data collection mentioned previously may have been exacerbated by the requirement for participants to focus on their feelings. An unwillingness or failure to pay attention to, and therefore maintain, emotional states, will result in poorer accuracy scores. However, several prior studies have found discrepancies between emotion and brightness maintenance task performance (Mikels et al., 2005, Mikels et al., 2008; Broome et al., 2012), and accuracy for both tasks in the current study are still well-within the range of previously observed performance. Furthermore, the present findings replicate and extend a previously identified relationship that was found when brightness and emotion maintenance task performance was both matched (Frank et al. Study 3, 2021) and unmatched (Frank et al. Studies 1 & 2, 2021). Therefore, we do not believe differences in performance on the maintenance tasks can account for the current findings.

Additionally, our sample size (N = 76) fell short of our goal (N = 85) by nine participants. This was likely a result of high rates of attrition (25%), intensified by the use of online data collection, as well as the Covid-19 global pandemic. Moreover, because the study timeline was bound to a national event—the 2020 U.S. election—we were unable to add additional participants after data collection had begun. Nevertheless, we had sufficient power to detect small to medium effects (Description-Based: Cohen's f = 0.123, Election Outcome: Cohen's f = 0.135) that supported our a priori hypotheses that affective working memory is related to the ability to predict future feelings. Furthermore, our sample falls within the large range of Ns recruited in previous forecasting studies that focused on an election outcome (approximately 60 – 700 participants; Gilbert et al. Study 3, 1998; Lench et al., 2019), albeit on the lower end of the spectrum.

Conclusion

In conclusion, we provide novel evidence that affective working memory supports forecasting abilities—not only in the laboratory, but in the real-world as well. The present work further establishes the importance of affective working memory as a fundamental ability that, like other forms of working memory, varies across individuals, and plays a key role in higher-order psychological processes, such as emotional prospection. The findings also provide a better understanding of mechanisms underlying variability in affective forecasting accuracy and reveal possible directions for improving these abilities.

References

Ahn, W.-Y., Rass, O., Shin, Y.-W., Busemeyer, J. R., Brown, J. W., & O'Donnell, B. F. (2012). Emotion-based reinforcement learning. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*(34), 78–83.

Broome, R., Gard, D. E., & Mikels, J. A. (2012). Test–retest reliability of an emotion maintenance task. *Cognition and Emotion, 26*, 737–747. http://dx.doi.org/10.1080/02699931.2011.613916

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin, 27*, 1480–1493. http://dx.doi.org/10.1177/01461672012711009

Charpentier, C. J., De Neve, J., Li, X., Rosier, J. P., & Sharot, T. (2016). Models of affective decision making: How do feelings predict choice? *Psychological Science, 27*(6), 763–775. https://doi.org/10.1177/0956797616634654

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34,* 819B.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Davidson, R. J., & Irwin, W. M. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences*, *3*, 11–21. https://doi.org/10.1016/S1364-6613(98)01265-0

DeFraine, W. C. (2016). Differential effects of cognitive load on emotion: Emotion maintenance versus passive experience. *Emotion*, *16*(4). 459–467. https://doi.org/10.1037/emo0000140

de Voogd, E. L., Wiers, R. W., Zwitser, R. J., & Salemink, E. (2016). Emotional working memory training as an online intervention for adolescent anxiety and depression: A randomised controlled trial. *Australian Journal of Psychology, 68,* 228–238. https://doi.org/10.1111/ajpy.12134

Dunn, E. W., Brackett, M. A., Ashton-James, C., Schneiderman, E., & Salovey, P. (2007). On emotionally intelligent time travel: Individual differences in affective forecasting ability. *Personality and Social Psychology Bulletin, 33,* 85–93. https://doi.org/10.1177/0146167206294201

Emanuel, A. S., Updegraff, J. A., Kalmbach, D. A., & Ciesla, J. A. (2010). The role of mindfulness facets in affective forecasting. *Personality and Individual Differences, 49*(7), 815–818. https://doi.org/10.1016/j.paid.2010.06.012

Frank, C. C., Iordan, A. D., Ballouz, T. L., Mikels, J. A., & Reuter-Lorenz, P. A. (2021). Affective forecasting: A selective relationship with working memory for emotion. *Journal of Experimental Psychology: General*. *150*(1), 67–82. https://doi.org/10.1037/xge0000780

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science, 317* (5843), 1351–1354. https://doi.org/10.1126/science.1144161

Hayes, W. M. & Wedell, D. H. (2020). Modeling the role of feelings in the Iowa Gambling Task. *Decision, 7*(1), 67–89. https://doi.org/10.1037/dec0000116

Hoerger, M., Chapman, B., & Duberstein, P. (2016). Realistic affective forecasting: The role of personality. *Cognition and Emotion, 30*, 1304–1316. https://doi.org/10.1080/02699931.2015.1061481

Hoerger, M., Chapman, B. P., Epstein, R. M., & Duberstein, P. R. (2012). Emotional intelligence: A theoretical framework for individual differences in affective forecasting. *Emotion, 12*, 716–725. https://doi.org/10.1037/a0026724

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science, 17*(6), 649–653. https://doi.org/10.1111/j.1467-9280.2006.01760.x

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39–58.

Lench, H. C., Levine, L. J., Perez, K., Carpenter, Z. K., Carlson, S. J., Bench, S. W., & Wan, Y. (2019). When and why people misestimate future feelings: Identifying strengths and weaknesses in affective forecasting. *Journal of Personality and Social Psychology, 116*(5), 724–742. https://doi.org/10.1037/pspa0000143

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology, 103*(4), 584–605. https://doi.org/10.1037/a0029544

Levine, L. J., Safer, M. A., & Lench, H. C. (2006). Remembering and misremembering emotions. In L. J. Sanna & E. C. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors* (pp. 271–290). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195177664.003.0015

Loewenstein, G. (2005). Hot-cold empathy gaps in medical decision making. *Health Psychology, 24*(4), S49–S56. https://doi.org/10.1037/0278-6133.24.4.S49

Loewenstein, G., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Loewenstein, G. & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, *118*(4), 1209–1248. https://doi.org/10.1162/003355303322552784

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281. https://doi.org/10.1038/36846

MacLeod, A. K. (2016). Prospection, well-being and memory. *Memory Studies*, *9*(3), 266–274. https://doi.org/10.1177/1750698016645233

Marroquín, B. & Nolen-Hoeksema, S. (2015). Event prediction and affective forecasting in depressive cognition: Using emotion as information about the future. *Journal of Social and Clinical Psychology, 34*(2), 117–134. https://doi.org/10.1521/jscp.2015.34.2.117

Mellers, B. A., Schwartz, A., & Ritov, I (1999). Emotion-based choice. *Journal of Experimental Psychology: General, 128*(3), 332–345. https://doi.org/10.1037/0096-3445.128.3.332

Mellers, B. A. & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10*(6), 210–214. http://dx.doi.org/10.1111/1467-8721.00151

Meyvis, T., Ratner, R. K., & Levav, J. (2010). Why don't we learn to accurately forecast feelings? How misremembering our predictions blinds us to past forecasting errors. *Journal of Experimental Psychology: General, 139*(4), 579–589. https://doi.org/10.1037/a0020285

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging*, *20*, 542–553. https://doi.org/10.1037/0882-7974.20.4.542

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion, 8*(2), 256–266. https://doi.org/10.1037/1528-3542.8.2.256

Mikels, J. A., & Reuter-Lorenz, P. A. (2013). Emotion and working memory. In H. Pashler (Ed.), *Encyclopedia of the mind* (pp. 308–310). Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781452257044.n113

Mikels, J. A., & Reuter-Lorenz, P. A. (2019). Affective working memory: An integrative psychological construct. *Perspectives on Psychological Science, 14,* 543–559. https://doi.org/10.1177/1745691619837597

Miranda, R., & Mennin, D. S. (2007). Depression, generalized anxiety disorder, and certainty in pessimistic predictions about the future. *Cognitive Therapy and Research, 31*(1), 71–82. https://doi.org/10.1007/s10608-006-9063-4

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2020). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology*, 0(0), 1–23. https://doi.org/10.1111/apps.12302

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 5,* 191–203. https://doi.org/10.3758/s13428-018-01193-y

Prajapati, B., Dunne, M., & Armstrong, R. (2010). Sample size estimation and statistical power analyses. *Optometry Today, 16*(7), 10–18.

Redick, T. S. (2019). The hype cycle of working memory training. *Current Directions in Psychological Science, 28,* 423–429. https://doi.org/10.1177/0963721419848668

Roepke, A. M. & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology, 55*, 23–48. https://doi.org/10.1111/bjc.12087

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review, 24,* 1077–1096. https://doi.org/10.3758/s13423-016-1217-0

Smith, R., Lane, R. D., Alkozei, A., Bao, J., Smith, C., Sanova, A., ... & Killgore, W. D. (2017). Maintaining the feelings of others in working memory is associated with activation of the left anterior insula and left frontal-parietal control network. *Social Cognitive and Affective Neuroscience*, *12*(5), 848–860. https://doi.org/10.1093/scan/nsx011

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy, 44*(6), 861–882. https://doi.org/10.1016/j.brat.2005.07.001

Wilson, T. D. & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35,* 345–411. https://doi.org/10.1016/S0065-2601(03)01006-2

Wilson, T. D., & Gilbert D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science, 14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology, 105*(5), 740–748. https://doi.org/10.1037/a0032662

Waugh, C. E., Lemus, M. G., & Gotlib, I. H. (2014). The role of the medial frontal cortex in the maintenance of emotional states. *Social Cognitive and Affective Neuroscience*, *9*(12), 2001–2009. https://doi.org/10.1093/scan/nsu011

Chapter 4 General Discussion

This dissertation aimed to determine whether affective working memory—the distinct domain of working memory responsible for the maintenance of emotional experiences—plays a fundamental role in affective forecasting. We hypothesized that individuals who are better able to hold emotions in mind (i.e., superior affective working memory) would be more accurate in their predictions of future feelings. In support of this prediction, the results from four studies (two pre-registered) demonstrate that individual differences in emotion, but not perceptual, maintenance performance predict forecasting accuracy. A specific role of affective working memory is further established by including additional measures of visuospatial working memory, which do not reliably predict forecasting accuracy (Study 3). Furthermore, this significant association holds true even when forecasting feelings about a major real-world event (Study 4). Below, we summarize the evidence and discuss the implications, future directions, and limitations of the work.

Summary of Findings

Across the four studies in this dissertation, we establish a novel and robust relationship between affective working memory and affective forecasting accuracy. In Study 1a (N = 66), we found that emotion maintenance predicts individual variability in affective forecasting. This association was further replicated across Studies 2 (N = 96) and 3 (N = 85). In Study 4 (N = 76), we find that this relationship between affective working memory and affective forecasting continues to hold, even when using a real-world measure of forecasting (i.e., 2020 U.S. Presidential Election) tested on-line across a broad sample of internet-based participants. Taken together, we present strong evidence, based on a total of 323 participants, that the ability to maintain emotional experiences supports accurate affective forecasting.

In comparison, there was no evidence to indicate that perceptual working memory can account for variability in forecasting abilities. Across all studies, we find that brightness-maintenance performance is unrelated to forecasting accuracy. This was also the case for two additional, more widely used measures of visuospatial working memory, which did not independently predict forecasting accuracy (Study 3). An additional analysis[7] suggested after publication further supports that non-affective working memory, operationalized as a composite of visuospatial working memory measures, does not predict forecasting accuracy. In addition, brightness maintenance did not predict accuracy of forecasted feelings about the U.S. election (Study 4). The lack of a relationship between cognitive working memory and forecasting accuracy underscores that the ability to actively maintain the intensity of an emotional experience may be a core component of the ability to predict future feelings.

Therefore, a consistent relationship emerges between forecasting accuracy and emotion, but not brightness, maintenance abilities. Although the standardized regression coefficients (i.e., beta weights; Frost, 2019) do not significantly differ between brightness-maintenance and emotion-maintenance predictors (see Appendix 4A for beta weight comparison), the identical pattern is evident across all studies such that emotion-maintenance accuracy significantly predicts forecasting accuracy, whereas brightness-maintenance does not (i.e., the slope does not

[7] Using a composite-based Structural Equation Modeling (SEM) approach, a visuospatial working memory composite score (comprised of Corsi block-tapping forward, Corsi block-tapping backward and change detection Z-scores) did not significantly predict forecasting accuracy, $\beta = -0.043$, $p = 0.782$. This was also the case when using a non-affective working memory composite score, which included the measures mentioned above, as well as brightness-maintenance performance, $\beta = -0.066$, $p = 0.695$.

differ from zero). The selective nature of this relationship with affective forecasting further establishes the separability of affective and non-affective working memory abilities.

A potential alternative explanation of these divergent associations with forecasting accuracy is that they stem from differences in the psychometric properties (i.e., reliability) of the emotion and brightness intensity rating tasks. If emotion and brightness intensity ratings differ in their intra-individual consistency over time, then the two maintenance tasks, which use these ratings to determine accuracy, might be differentially sensitive to individual differences in performance. As a result, relationships with other constructs, such as affective forecasting, may also differ. However, in an independent assessment of rating reliability (see Appendix 4B for reliability analyses), we found no differences in reliability estimates between emotion and brightness intensity ratings. This outcome aligns with our interpretation that the selective association we've observed reflects the unique role that affective working memory plays in emotional prospection.

Another possibility is that participants are performing the emotion maintenance task differently than instructed. That is, instead of maintaining the feelings evoked by the photographs, participants could be verbally recoding the emotional intensity of each image as a numeric code to remember. If this were the case, then performance on the emotion maintenance task would reflect more verbal, rather than affective, working memory abilities. However, Mikels and colleagues (2008) addressed this potential explanation by having participants engage in articulatory suppression while doing the emotion maintenance task at the same time. If participants are using a verbal strategy to complete the emotion maintenance task, concurrent articulatory suppression should interfere with verbal recoding and impair accuracy. Yet, Mikels et al. (2008) found that simultaneous performance of these tasks actually improved, not worsened, emotion-maintenance accuracy. Additionally, in Studies 1–3 of the present dissertation, participants were required to pass an "instruction comprehension" check before beginning each maintenance task and again, once the experiment had ended. Thus, although one can never be certain, it appears improbable that participants ignore the stated instructions of the emotion maintenance task in favor of alternative strategies.

Additionally, across the first three studies, we examined relationships between forecasting accuracy and a number of exploratory measures including visual imagery, emotional intelligence, and emotion regulation. We did not find a relationship between forecasting accuracy and visual imagery (Studies 1, 2, 3), nor between forecasting accuracy and trait or ability emotional intelligence (Studies 1 & 2). However, we did find initial evidence that better forecasting accuracy may be related to more frequent cognitive reappraisal—an emotion regulation strategy that involves the mental reinterpretation of emotional stimuli (Study 3; Gross, Richards, & John, 2006). Thus, further examination of these relationships with affective forecasting, especially cognitive reappraisal, will lead to a broader, and more thorough, understanding of how individuals accurately predict their feelings.

Implications & Future Directions

While affective working memory has previously been studied in variety of contexts, the present studies, to our knowledge, are among the first to examine how individual differences in working memory for emotion contribute to another mental ability. Evidence from this dissertation reveals a direct association between affective working memory and an important form of prospective thought. These findings support the idea that affective working memory is a fundamental psychological process and may support other forms of higher-order emotional abilities. According to this formulation, just as cognitive working memory provides the mental workspace for complex cognitive tasks, affective working memory provides the mental workspace for mental simulations where feelings play a prominent role. However, affective forecasting is unlikely to be the only process using such a workspace. Thus, future research should aim to identify other emotion-related abilities supported by affective working memory. It may be case that affective working memory plays a similar role in other types of emotional prospection (e.g., counterfactual thinking), or emotion-based choice more generally.

This work also has important implications for affective forecasting. While inaccurate prospection (i.e., future thinking), including affective predictions, is associated with suboptimal decision making and negative mental health and well-being outcomes (Charpentier et al., 2016; Gilbert & Wilson, 2007; MacLeod, 2016; Roepke & Seligman, 2016), previous studies have found that prospection can be improved through the use of interventions (Buehler & McFarland, 2001; Roepke & Seligman, 2016). The evidence that affective working memory supports accurate forecasting indicates that an intervention specifically targeting emotion-maintenance abilities may be particularly effective in enhancing forecasting accuracy and related processes. This may be best achieved through the use of training, which has been found to improve working memory performance in the cognitive domain (e.g., Soveri et al., 2017; see also Redick, 2019), and may also be successful in the affective domain (e.g., de Voogd et al., 2016). However, given the current debate on whether working memory training benefits are successfully transferred to other, related abilities (Redick, 2019), further research is needed to establish whether augmenting affective working memory abilities will lead to improved forecasting accuracy. Nevertheless, if found to be successful, these programs may be especially beneficial for certain clinical populations—such as individuals with Schizophrenia or depression—who have documented impairments in emotional future-thinking abilities (Gard et al., 2007; Hoerger et al., 2012; Wenze, Gunthert & German, 2012).

Limitations

One methodological limitation of this work is that the design of our core maintenance tasks is not optimized to determine the capacity of affective working memory per se—that is, the maximum number of emotional experiences participants can hold in mind at the same time. In the current emotion maintenance task, participants hold feelings elicited from one image in mind over a delay, before comparing the intensity of their feelings to emotions elicited by a second image. Thus, for each trial, participants are primarily responsible for maintaining one emotional experience. This is also true of the brightness maintenance task, which requires the maintenance of one subjective experience of brightness. However, there is robust evidence that people are limited in how many items they can hold in mind at one time, at least for non-affective information (Cowan, 2010). Although not yet tested, a similar boundary is likely to exist for the working memory for emotion such that the number of emotional experiences people are able to hold in mind is also constrained. Supporting evidence comes from the finding that concurrent performance of an emotion maintenance and emotion regulation task, which requires the focus of another, separate emotional experience, impairs affective working memory abilities (Mikels et al., 2008). We also know that estimates of cognitive working memory capacity vary across individuals with some people able to retain more information than others. We believe that similar variability would be found in affective working memory capacity, with some individuals able to maintain more emotional experiences than others. Alternatively, the limits of affective working memory may also be related to the precision or duration of maintaining emotional experiences. Therefore, while further investigation into the boundaries of affective working memory was outside of the scope of the current dissertation, we recognize that doing so in the future would allow for a more comprehensive understanding of affective working memory as a construct, and provide additional insight into the role it plays in affective forecasting and other types of higher-order emotional thought.

Furthermore, a caveat to the finding that cognitive working memory fails to predict individual differences in forecasting accuracy is that we did not assess verbal working memory abilities. Across all four studies presented here, non-affective working memory was measured using the brightness maintenance task, a non-emotional companion measure designed to match the demands of the emotion maintenance task (i.e., maintaining the intensity of a subjective experience elicited by images taken from the International Affective Picture System (IAPS; Lang, Bradley and Cuthbert, 1997). Additionally, in Study 3, we included two widely used measures of visuospatial working memory, Corsi Block Tapping (Corsi, 1972) and a visual change detection task (Luck & Vogel, 1997), selected because their perceptual focus made them suitable for comparison to the brightness maintenance task. Hence, because we did not include a measure of verbal working memory, we cannot discount the possibility that working memory for verbal information would contribute to forecasting accuracy. However, Kane et al. (2004) finds that across several tasks, verbal and visuospatial working memory abilities are related to one another. Additionally, one study that used operation and reading span tasks (see Conway et al., 2005 for a review on working memory measures) to assess verbal working memory abilities, found that working memory capacity was only marginally (p = .06) related to an overestimation forecasting bias, and only for a negative, but not positive, outcome (Hoerger et al., 2010). Moreover, this association was assessed in an intervention study where half of the participants had undergone a program designed to reduce forecasting bias, thereby making these findings unlikely to generalize to other contexts. As a result, we believe the failure of cognitive working memory to account for variability in forecasting accuracy would be evident even when tested using other measures of working memory, including verbal tasks. Nonetheless, this is an empirical question that awaits future research.

Closing Remarks

Evidence from four studies presented in this dissertation identifies affective working memory as a core ability underlying the accuracy of emotional prospection. We find that the predictive association with affective forecasting is selective to the working memory for emotion, and that this association extends to forecasted feelings about a major real-world event. These findings provide further support for the unique role of affective working memory as a fundamental psychological process supporting higher-order emotional thought. Moreover, these results suggest that affective working memory may be a promising target for interventions designed to improve our capacity to maintain emotional experiences, which can in turn, improve people's ability to anticipate their future emotions and make better decisions.

References

Charpentier, C. J., De Neve, J. E., Li, X., Roiser, J. P., & Sharot, T. (2016). Models of affective decision making: How do feelings predict choice? *Psychological Science*, *27*(6), 763–775. https://doi.org/10.1177/0956797616634654

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Dissertation Abstracts International, 34, 819B.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51–57. https://doi.org/10.1177/0963721409359277

de Voogd, E. L., Wiers, R. W., Zwitser, R. J., & Salemink, E. (2016). Emotional working memory training as an online intervention for adolescent anxiety and depression: A randomised controlled trial. *Australian Journal of Psychology*, *68*(3), 228–238. https://doi.org/10.1111/ajpy.12134

Frost, J. (2019). *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models.*

Gard, D. E., Kring, A. M., Gard, M. G., Horan, W. P., & Green, M. F. (2007). Anhedonia in schizophrenia: Distinctions between anticipatory and consummatory pleasure. *Schizophrenia Research*, *93*(1–3), 253–260. https://doi.org/10.1016/j.schres.2007.03.008

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the Future. *Science, 317*(5843), 1351–1355. https://doi.org/10.1126/science.1144161

Gross, J. J., Richards, J. M., & John, O. P. (2006). Emotion regulation in everyday life. In D. K. Snyder, J. Simpson, & J. N. Hughes (Eds.), *Emotion regulation in couples and families: Pathways to dysfunction and health* (pp. 13–35). American Psychological Association. https://doi.org/10.2307/2076895

Hoerger, M., Quirk, S. W., Chapman, B. P., & Duberstein, P. R. (2012). Affective forecasting and self-rated symptoms of depression, anxiety, and hypomania: Evidence for a dysphoric forecasting bias. *Cognition and Emotion, 26*(6), 1098–1106. https://doi.org/10.1080/02699931.2011.631985

Hoerger, M., Quirk, S. W., Lucas, R. E., & Carr, T. H. (2010). Cognitive determinants of affective forecasting errors. *Judgment and Decision Making, 5*(5), 365–373.

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research, 25*(4), 467-478. https://doi.org/10.1207/s15327906mbr2504_4

Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention,* 39–58.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. https://doi.org/10.1038/36846

MacLeod, A. K. (2016). Prospection, well-being and memory. *Memory Studies, 9*(3), 266–274. https://doi.org/10.1177/1750698016645233

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion, 8*(2), 256–266. https://doi.org/10.1037/1528-3542.8.2.256

Redick, T. S. (2019). The hype cycle of working memory training. Current Directions in *Psychological Science, 28*(5), 423–429. https://doi.org/10.1177/0963721419848668

Roepke, A. M., & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology, 55*(1), 23–48. https://doi.org/10.1111/bjc.12087

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review, 24*(4), 1077–1096. https://doi.org/10.3758/s13423-016-1217-0

Wenze, S. J., Gunthert, K. C., & German, R. E. (2012). Biases in affective forecasting and Recall in individuals with depression and anxiety symptoms. *Personality and Social Psychology Bulletin, 38*(7), 895–906.

Appendices

Appendix 2A: Affective Forecasting Stimuli

Table 2A- 1. Affective Forecasting Stimuli

IAPS Image	Description
1750	This scene shows two bunny rabbits sharing a stalk of grass together
6230	This scene shows a gun pointed straight at you. It appears to be inches away from your face.
6315	This scene shows a teenage girl reeling in pain having just been hit by a man whose hand is clenching her throat.
1304	This scene shows a wild wolf-like animal directly in front of you. It has its jaws wide open and is apparently ready to attack.
8467	This scene shows world-class runners sprinting for the finish line in what looks like a very important race
4660	This scene shows two young lovers who are kissing each other passionately. They are attractive and, from what you can see, naked.
3170	This scene shows a baby with a facial tumor. The tumor, which is protruding through an eye socket, is also causing one side of the infant's head to bulge out.
7580	This scene shows a desert with gently rolling sand dunes, a setting sun, and a deep blue sky. The desert is very still.
8496	This scene shows children barreling down a water slide. They are yelling for joy and having a great time
9301	This scene shows a filthy toilet that hasn't been cleaned in a very long time. The toilet is full of feces and other waste products.

Appendix 2B: Analyses of Directionality in Affective Forecasting Errors

Collapsing across participants included in analyses from Studies 1a, 1b, 2, and 3 (N = 315), the average difference between predicted and experienced feelings was M = –0.11 (SD = 1.05; raw score range: –10 to +10), indicating no overall bias for participants to overestimate or underestimate how intensely they would feel, $t(314) = -1.92$, $p = .056$, $d = .11$. However, we found a significant difference in deviation scores between positive (M = –1.11, SD = 1.74) and negative (M = .89, SD = 1.52) forecasts such that there was a tendency for participants to overestimate how negatively they would feel for negative trials but underestimate how positively they would feel on positive trials, $t(314) = 14.16$, $p < .001$, $d = 1.22$. As can be seen in Table 2B-1, this pattern is evidenced in each individual study as well.

Table 2B- 1. Mean (SD) Forecasting Bias by Valence Across Studies 1–3

Valence	Study 1a N = 66	Study 1b N = 68	Study 2 N = 96	Study 3 N = 85
Negative	0.98 (1.53)	1.06 (1.74)	0.77 (1.38)	0.81 (1.5)
Positive	–0.87 (1.84)	–1.45 (2.03)	–1.0 (1.67)	–1.17 (1.46)

Range of Deviation: –10 to +10 where negative values reflect underestimation and positive values reflect overestimation.

Appendix 2C: Pair Selection using Item Response Theory

Based on the data obtained from Studies 1a and 1b, item response theory (IRT) was used to select the subsets of stimuli for each maintenance task to be included in Studies 2 and 3. We used a series of 2-parameter logistic models, which allows for estimation of the threshold, or the difficulty of the item, and the slope, or how related the item is to the construct being measured by the task (Harris, 1989). The slope estimate is also referred to as "item discrimination", which is the criteria we used to determine which pairs were most related to the construct being measured. Eligible pairs (see below) were entered into the models and those with the highest discrimination estimates were selected for inclusion in the refined stimulus sets. The general formula for the 2-parameter logistical model we used in the present analyses is

$$P(y_{ij}=1|\theta, \delta, \alpha) = \text{logis}(\alpha_j(\theta_i - \delta_j))$$

where the probability of getting an item correct, P, is a function of the difficulty of the item (δ_j), the ability (θ_i) of the individual, *i*, and the discrimination of the item, α_j.

Affect Maintenance. Trials from all participants in Study 1a were coded as correct or incorrect to be included in one of two IRT models, performed separately for positive and negative pairs.

Brightness Maintenance. After exclusion of images due to content as described in the main text, remaining trials were coded as correct or incorrect for inclusion in the model. This analysis was based on the subset of participants in Study 1b who performed the task in the same image order as was to be used in subsequent studies.

Study 2

For the affect maintenance task, the 28 pairs with the highest discrimination estimates for each valence from Study 1a were selected for inclusion for Study 2, making up 56 pairs total. For the brightness maintenance task, the 28 pairs with the highest discrimination scores from Study 1b were selected for inclusion in Study 2.

Study 3

Noting that the overall accuracy of the two maintenance tasks differed significantly in Study 2, we reconsidered our approach for selecting the most diagnostic pairs to achieve more matched performance. Accordingly, for Study 3, the most discriminating pairs were selected for each task, with the constraint that the discrimination estimates would be matched between the 28 brightness maintenance (M = .69, SD = .40), and 56 affect maintenance (M = .67, SD = 0.34) pairs, $t(82) = .19$, $p = .847$. To further verify the approximate equivalence of these estimates, a Bayesian independent samples t-test using JASP (Version 0.9) was also calculated with Cauchy priors (scale = .707) to determine the Bayes Factor (BF_{01}) score, which compares the likelihood of the null and alternative hypotheses given the data. A BF_{01} of 4.11 revealed substantial support for no difference in discrimination estimates between image sets, indicating the data were more than four times more likely to be explained by the null hypothesis. Additionally, within the affect maintenance stimuli, the discrimination estimates were matched between the 28 positive (M = 0.66, SD = 0.35) and 28 negative (M = 0.68, SD = 0.34) pairs, $t(54) = .23$, $p = .82$. A Bayesian independent samples t-test revealed a BF_{01} of 3.65, indicating strong support for the null hypothesis that there is no difference between discrimination estimates between positive and negative image sets.

References

Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, *8*(1), 35-41. http://dx.doi.org/10.1111/j.17453992.1989.tb00313.x

JASP team. JASP (Version 0.9) [Computer software]. Retrieved from https://jasp-stats.org/previous-versions/

Appendix 3A: Life Events Forecasting Measure

This study also included a novel, exploratory measure of event-based affective forecasting that compared predicted and experienced feelings to everyday life events. The majority of existing event-based forecasting measures use major national public (e.g., election) or personal (e.g., exam performance, Valentine's day) events that rely on a predictable but constrained timeline. Our aim was to create a forecasting measure that incorporated a range of potential personal, real-life experiences, as opposed to a single target event, in an effort to eliminate some of the scheduling constraints inherent in other measures, and to develop a measure intended to be more representative of the kind of forecasting individuals engage in on a daily basis. We adapted an event prediction measure used originally by Marroquín and Nolen-Hoeksema (2015) to examine event likelihood and predicted feelings about potential events. Unlike Marroquín and Nolen-Hoeksema however, we include a follow-up assessment where we also collect ratings of experienced feelings *after* the events have occurred. Obtaining both predicted and experienced feelings allows for the calculation of the discrepancy, and therefore, the accuracy, of individuals' forecasts.

To create this measure, we ran a pilot study to select items from a pool of 96 life events—84 taken from Strunk, Lopez & DuRubeis (2006), Miranda & Mennin (2007), and Marroquín & Nolen-Hoeksema (2015), and 12 original items we created to be relevant to current events (e.g., the Covid-19 pandemic). Participants in this pilot study (N = 77) rated the likelihood and predicted feelings to each of the 96 potential events. After a 3-week delay, participants reported which events occurred and rated their experienced feelings for each item. To maximize the likelihood that future participants would experience the events we included in the new measure, we selected 28 (14 positive, 14 negative) of the most commonly experienced life events (i.e., items with the highest reported frequency among pilot participants), while also maintaining an equal number of positive and negative items, matched in likelihood and predicted feeling intensity. Ratings of experienced feelings were used to determine forecasting accuracy exclusion criteria (i.e., outliers) for the primary study, but were not used for item selection.

Method

We created a novel measure of affective forecasting by modifying an event prediction task (Marroquín & Nolan-Hoeksema, 2015) to compare predicted and experienced feelings to life events. In Phase I (Session 1), participants predicted how they would feel about an event (e.g., you have a supervisor or teacher praise your work; see Table 3A-1 for full list of items) if it were to happen in the near future using a scale (absolute resolution: 21 points) from *not at all happy* to *extremely happy*. Three weeks later, in Phase II (Session 3), participants reported how they felt about each event that happened in the moment it occurred using the same scale. Participants also reported when the event had occurred (in approximate weeks) and how they felt about each event 'right now,' that is, at the time of rating. Participants reported experiencing an average of 9.33 (4.52 positive, 4.81 negative; see Table 3A-1 for the reported frequencies of each event) of the 28 possible events, that took place approximately 1 week prior to rating (M = 1.01 weeks, SD = .99). Error scores were calculated as the absolute difference between Phase I and Phase II ratings (possible range: 0 – 20; score >10 indicates valence reversal), averaged across all items that occurred for that participant. Two outliers were removed (forecasting error > 3 SD above the mean), and three additional participants did not report experiencing any of the 28 events.

Relationships with Maintenance Task Performance

As outlined in our preregistration (https://aspredicted.org/blind.php?x=et4i2p), we used a multivariate linear regression to assess whether accuracy on the life events forecasting measure was significantly predicted by emotion-maintenance and brightness-maintenance performance. As expected, brightness-maintenance performance did not predict forecasting accuracy, $\beta = .163$, $p = .210$. Yet, unlike the other measures of forecasting abilities, emotion-maintenance performance did not account for individual variability in forecasting on the life events measure, $\beta = –.11$, $p = .393$. The absence of a significant relationship between emotion maintenance and forecasting accuracy on the life events measure may be due to several unique features of this task. First, in the majority of studies that examine affective forecasting, accuracy scores are calculated using the same specific events and approximate timing for all participants. In contrast, for this life events measure, forecasting accuracy was calculated based only on the events that happened to each participant and the specific details of those events could vary. This led to discrepancies between which, and how many, events were used to determine each participant's forecasting abilities, adding additional sources of variance to these scores. Furthermore, in the life events measure, accuracy was based on participants' retrospective ratings of how they felt when the event occurred. Consequently, the time between event occurrence and rating varied among events for a given individual, and among individuals. Furthermore, people had to rely on long-term memory to report rate their feelings. Because memory for our previous emotions is subject to considerable error (see Levine, Safer, & Lench, 2006 for review), using past, instead of present, ratings of feelings may have added further variability to this measure, thereby potentially obscuring any relationship with emotion-maintenance ability. Thus, future work is necessary for determining if and how this measure can be used in future forecasting experiments. It may be particularly useful to explore the unique characteristics of the life events (e.g., expectedness, frequency, overall relevance) to determine which items are most sensitive to forecasting error, and therefore, are most suitable for creating a composite forecasting score.

Relationships with Other Forecasting Measures

We also directly compared forecasting on the life events task to performance on the other forecasting measures used in this study. We found that participants were more accurate in their forecasted feelings to life events (M = 2.29, SD = 1.15) than for the election outcome (M = 2.81, SD = 2.22, $t(78) = -2.05$, p = .044), and marginally more accurate compared to description-based forecasting (M = 2.59, SD = 1.04, $t(80) = 1.88$, p = .06). Nevertheless, forecasting error on the life events measure was related to error on both the description-based ($r(79) = .228$, p = .041) and election outcome ($r(77) = .321$, p = .004) measures, although the association between description-based and election outcome forecasting error was not significant, $r(79) = .13$, p = .248). These relationships between the life events and other forecasting measures may stem from a common component of forecasting other than affective working memory, possibly related to other aspects of introspection. Thus, while follow-up work aimed at obtaining a better understanding of this task is required, these significant associations indicate there may be some promise for this method to measure forecasting accuracy in future studies.

References

Levine, L. J., Safer, M. A., & Lench, H. C. (2006). Remembering and misremembering emotions. In L. J. Sanna & E. C. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors* (pp. 271–290). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195177664.003.0015

Marroquín, B., & Nolen-Hoeksema, S. (2015). Event prediction and affective forecasting in depressive cognition: Using emotion as information about the future. *Journal of Social and Clinical Psychology, 34*(2), 117–134. https://doi.org/10.1521/jscp.2015.34.2.117

Miranda, R., & Mennin, D. S. (2007). Depression, generalized anxiety disorder, and certainty in pessimistic predictions about the future. *Cognitive Therapy and Research, 31*(1), 71–82. https://doi.org/10.1007/s10608-006-9063-4

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy, 44*(6), 861–882. https://doi.org/10.1016/j.brat.2005.07.001

Table 3A- 1. Life Events Forecasting Items and Frequency

Item	**Frequency Count**	**Proportion[1] of Sample that Reported Each Event**
You feel misunderstood by people	34	0.405
You experience a moment of great insight	30	0.357
Things don't work out as hoped	47	0.560
You go out of town for leisure	13	0.155
You successfully teach someone a new skill or concept	25	0.298
You receive a call from a telemarketer	27	0.321
You fall badly behind in work	18	0.214
You feel socially inadequate.	36	0.429
You go out to dinner and have a relaxing and enjoyable conversation with friends.	22	0.262
You are unable to cope with your responsibilities	23	0.274
You learn a new skill related to work or school	18	0.214
You are admired by people	27	0.321
You have an inspiring conversation	31	0.369
You read and complete a book	18	0.214
You have an unexpected expense	24	0.286
You have a supervisor or teacher praise your work	29	0.345
You often have work or school go smoothly	45	0.536
You have a headache	46	0.548
You are considered to be an excellent listener	23	0.274
You have lots of good times with friends	36	0.429
You feel unable to confide in anyone	25	0.298
You reach out to a good friend who lives far away	38	0.452
You have hard work acknowledged at work or in class	27	0.321
You see an offensive post on social media	37	0.440
An activity you were planning on attending is postponed due to the pandemic	25	0.298
You forget your mask and you have to return home to retrieve it	18	0.214
A non-masked stranger moves too close for comfort	33	0.393
You hear of another devastating wildfire	9	0.107

[1]Proportions were calculated using data from the 84 participants who completed both phases

Appendix 4A: Comparing Beta Weights of Maintenance Task Predictors

We assessed whether there was significant difference between how well emotion-maintenance and brightness-maintenance performance predicted forecasting accuracy by comparing their beta weights. Beta weights (β), or standardized regression coefficients, represent the effect of change in each predictor (i.e., maintenance task performance) on the outcome variable (i.e., forecasting accuracy), holding other predictors constant (Jaccard, Wan, & Turrisi, 1990; Frost, 2019). To test whether beta weights differed between emotion-maintenance and brightness-maintenance predictors, we ran an additional regression model for each study[8] that included a maintenance type (emotion, brightness) by maintenance accuracy interaction term as an additional predictor of forecasting accuracy (Jaccard et al., 1990; Frost, 2019). A significant interaction term would indicate that forecasting accuracy was differentially predicted by performance on the brightness and emotion maintenance tasks. Across Studies 2, 3, & 4, we did not find significant differences between emotion-maintenance and brightness-maintenance predictors (see Table 4A-1)[9]. However, a clear pattern emerged across all studies such that emotion maintenance significantly predicted variability in forecasting accuracy, whereas brightness maintenance did not.

References

Frost, J. (2019). Introduction to statistics: An intuitive guide. *Statistics by Jim publishing: State College, PA, USA*, 196-204.

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research, 25*(4), 467-478. https://doi.org/10.1207/s15327906mbr2504_4

---

[8] This analysis could not be performed for Study 1, where a between-subject design was employed such that participants completed either the emotion maintenance or brightness maintenance tasks.
[9] Results do not change when tested using a robust, rather than ordinary least squares, regression.

Table 4A- 1. Comparison of Beta Weights Between Brightness-Maintenance and Emotion-Maintenance Predictors

	Brightness-Maintenance Predictor		Emotion-Maintenance Predictor		Interaction Term	
	β	*p*	β	*p*	β	*p*
Study 1	–0.20	.102	–0.32	.008*	NA	NA
Study 2	–0.16	.103	–0.33	.001*	0.118	.418
Study 3	0.03	.802	–0.25	.022*	0.246	.103
Study 4: Description-Based Forecasting	–0.08	.503	–0.29	.023*	0.072	.693
Study 4: Election Outcome Forecasting	0.18	.149	–0.33	.009*	0.300	.097

*$p < .05$
Note: β = Beta weight, or standardized regression coefficient

Appendix 4B: Test-Retest Reliability of Emotion and Brightness Intensity Ratings

Across the four studies in the present dissertation, emotion, but not brightness, maintenance task performance predicted affective forecasting accuracy, suggesting that affective working memory plays a unique role in forecasting future feelings. An alternative explanation is that this selective relationship is due to differences in the reliability of rating emotion versus brightness intensity of images used in the present research. Upon completion of each maintenance task, participants rate the emotion or brightness intensity of the corresponding images as they appear individually and in a random order. These ratings are then used to determine individualized accuracy on the maintenance tasks by inferring how participants should have responded during each maintenance trial. Therefore, if there are differences between the consistency of how individuals rate emotion, compared to brightness, intensity, then using these ratings as the basis for calculating maintenance task performance could be problematic. For instance, differences in the reliability of the intensity ratings may cause the emotion and brightness maintenance tasks to be differentially sensitive to individual differences in performance. If this were the case, then relationships identified with maintenance accuracy using these ratings may not reflect the true associations between constructs. Therefore, we computed and compared the test-retest reliability of the brightness intensity and emotion intensity ratings in an independent sample of participants.

Method

A total of 35 participants (ages 18 –29, mean (SD) age = 22.68 (3.3), 63% female, 71% students) completed the study. This sample size exceeded the goal of 30 participants, set in accordance with guidelines from Koo & Li (2016). Each participant completed two blocks of the emotion intensity rating task (112 images) and two blocks of the brightness intensity rating task

(56 images). During each trial, participants viewed an image and rated either the emotion intensity (i.e., strength or amount of emotional reaction) or brightness intensity (i.e., amount of perceived overall light or illumination) using a visual analog scale ranging from *Not At All Intense* to *Extremely Intense*. At the onset of each of the four blocks, participants were instructed which rating task (brightness intensity or emotion intensity) to complete and on-screen cues during each trial served as additional reminders. To minimize any potential order or sequence effects, block order was counterbalanced (i.e., ABBA, ABAB, BAAB, BABA) across participants. Participants were recruited via Prolific and completed the study online from their personal computers.

Results

To assess the test-retest reliability of the brightness intensity and emotion intensity rating tasks, we calculated intraclass correlation coefficients (ICC). ICC values reflect the proportion of variance accounted for by the random effects, where an estimate of zero reflects no reliability and an estimate of one reflects perfect reliability (Koo & Li, 2016). We calculated the ICC estimates and 95% confidence intervals based on the absolute agreement of ratings using a 2-way mixed model (fixed effect: time of measurement; random effects: participants, images), computed separately for brightness intensity and emotion intensity ratings. The ICC estimates for the brightness intensity and emotion intensity ratings were 0.444 (95% CI [0.297, 0.594]) and .447 (95% CI [0.331, 0.564]), respectively. As can be seen in Figure 4B-1, there is substantial overlap between the 95% confidence intervals of these estimates, indicating there is no difference between the reliability of emotion and brightness intensity ratings.

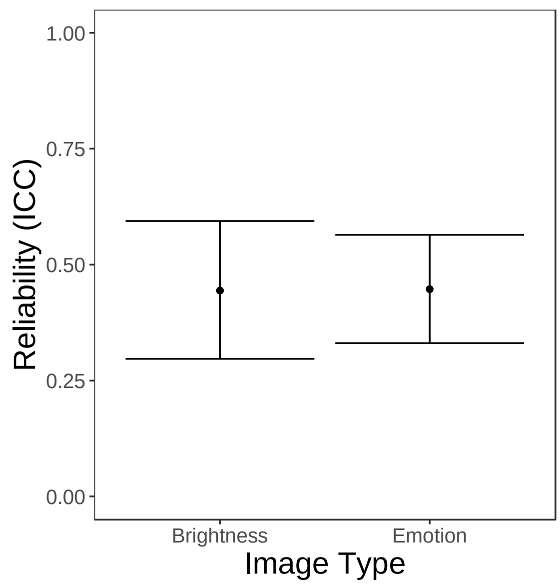

Figure 4B-1. Reliability estimates for brightness and emotion intensity ratings.

Discussion

We found no difference in test-retest reliability between ratings of emotion and brightness intensity—the ICCs were approximately equivalent. This suggests that differences in the consistency of ratings cannot account for the selective relationship between affective working memory and forecasting accuracy identified in this dissertation. Of note, the reliability estimates for both rating tasks are considered ‘poor,’ according to psychometric standards (Koo & Li, 2016). This is likely a result of the low between-subjects variability of the intensity ratings, which has been shown to reduce reliability (Hedge, Powell, & Sumner, 2017). Because these images were selected based on their ability to evoke emotional reaction (Bradley & Lang, 2007; Mikels et al. Pilot Studies A & B, 2008) or normative ratings of subjective experiences of brightness (Mikels et al. Pilot Study C, 2008), it is not surprising that intensity ratings would be similar across participants. And while the lack of variability in these ratings would be a concern if they were being directly tested for relationships with other constructs, these ratings are used solely for scoring the maintenance tasks—which, as seen across the present studies, do yield individual differences in performance. Furthermore, reliability for the maintenance tasks themselves has been assessed with evidence of significant test-retest reliability in maintenance abilities, with the exception of the most difficult brightness-maintenance trials (Broome, Mikels & Gard, 2012). Overall, while reliability of the intensity ratings may be limited, estimates are consistent between the emotion and brightness domains, and thus cannot explain differences between the associations with forecasting accuracy.

References

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. B. Allen (Eds.), *Series in affective science. Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford University Press.

Broome, R., Gard, D. E., & Mikels, J. A. (2012). Test-retest reliability of an emotion maintenance task. *Cognition and Emotion, 26*(4), 737–747. https://doi.org/10.1080/02699931.2011.613916

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion, 8*(2), 256–266. https://doi.org/10.1037/1528-3542.8.2.256